MONITORING RESERVOIR STORAGE OVER SOUTH ASIAN INTERNATIONAL RIVER BASINS USING REMOTE SENSING

A Dissertation

by

SHUAI ZHANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Huilin Gao
Committee Members,	Michael Bishop
	Andrew Klein
	Gretchen Miller
Head of Department,	Robin Autenrieth

August 2017

Major Subject: Civil Engineering

ABSTRACT

Satellite remote sensing has offered a unique promise for monitoring reservoir storage variations at near real-time. Such information is essential for flood mitigation—especially for regions dominated by international river basins like South Asia. In this dissertation, by using multi-satellite remote sensing observations, a series of algorithms was developed to improve the capability of monitoring reservoir storage variations at high temporal resolution and improved spatial coverage. These algorithms are presented in three studies.

The goal of the first study is to generate a first of its kind remotely sensed reservoir storage dataset for South Asia. Reservoir storage variations were inferred by combining water surface area (obtained by classifying optical satellite images) and elevation measurements (obtained from satellite laser altimeter measurements). This resulted in a 13-year dataset containing estimations for a total of 21 South Asian reservoirs, which represents 28% of the integrated reservoir capacity in the region. The storage estimates were highly correlated with observations—the coefficients of determination (R^2) were larger than 0.81, with a normalized root mean square error (NRMSE) ranging from 9.51% to 25.20%.

The second study explores the solution towards monitoring reservoirs at a high temporal resolution under all-weather conditions. Because optical satellite images suffer from cloud contamination during the rainy season, the developed remote sensing reservoir dataset can be restrained from providing the critical information necessary for

flood mitigation. A novel algorithm was developed by fusing passive microwave observations with optical satellite observations. This new algorithm has the advantage of working under all-weather conditions, and it reduces the reservoir monitoring time intervals from 10+ days to 4 days.

The third study further extends the spatial representation of the remotely sensed reservoirs from the first algorithm. Although the laser altimeter measurements are able to capture many reservoirs undetectable by traditional radar altimeters, they are still insufficient to form a dense observation network at a basin scale. In order to extend the spatial coverage, a new algorithm was developed to estimate reservoir storage by using water surface area values from MODIS imageries and surface elevation values from Digital Elevation Model (DEM) data (collected by the Shuttle Radar Topography Mission; SRTM). Using the SRTM based method, the spatial coverage of the South Asian reservoir dataset is extended from 28% to 45% of the overall storage capacity in South Asia.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Huilin Gao, for all of her support and guidance throughout these four years. She inspired me with great ideas and direction, but also gave me the freedom and independence necessary for conducting my research. Her steady guiding hand, wisdom, patience, and high standards have been invaluable to me and my development as a research scholar. My gratitude also goes out to my committee members—Dr. Michael Bishop, Dr. Andrew Klein, and Dr. Gretchen Miller—for their insightful suggestions and constructive criticism on my research.

I am grateful to my colleagues in Dr. Gao's research group for their help, support, and the great times they have brought me. I am so proud to have been a part of this inspiring team, and have greatly benefited from the weekly meetings and other interactions I have had with them. I would especially like to thank the organizations and individuals who provided data for my study, and the TAMU Supercomputing Facility for providing computing resources.

Last but not least, I would like to thank my wife, my parents, my sister, and my brother-in-law who have always supported my endeavors and given me the most essential elements for success − happiness and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Dr. Huilin Gao [advisor], Dr. Michael Bishop [Outside Department], Dr. Andrew Klein [Outside Department], and Dr. Gretchen Miller. An early advice for Chapter II was provided by David Kroodsma during the early stage of this work.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by startup funds from the Texas A&M University College of Engineering and the Zachry Department of Civil Engineering.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Effective reservoir operations can reduce flood inundation significantly [Mateo et al., 2014]. However, in many regions the sharing of storage information of reservoirs is limited (especially in developing countries), which critically hinders the capability of reservoir based flood mitigation [Adhikari et al., 2010]. Across the world, South Asia is probably the region that suffers the most from the negative impacts brought on by these limitations. Therefore, there is a strong societal need to advance the scientific understanding of the flood regimes in South Asia for reducing flood related losses.

Over the last few decades, the advent of satellite remote sensing has offered an unprecedented promise of monitoring lakes and reservoirs from space [Alsdorf et al., 2007; Lettenmaier et al., 2015]. Although the elevation levels of nearly 300 large lakes and reservoirs are observable globally using satellite radar altimeters, the current capability for monitoring South Asian reservoirs to assist flood mitigation is extremely limited, primarily due to two reasons:

First, few reservoirs in South Asia are observable by radar altimeters due to sensor limitations. For example, remotely sensed surface elevation values from radar altimeters are only available for 6 large reservoirs, which represent 10.7% of the capacity in the region. Furthermore, these radar altimetry measurements are only available sporadically, with an average continuous record of 18 years. Although ICESat/GLAS has a spatial resolution of 70 m—which makes it capable of capturing more reservoirs in South

Asia—its short lifetime (2003-2009) and long repeat period (91 days) have hindered its usage for monitoring purposes [Gao, 2015].

Second, the low temporal resolution (10+ days) of typical reservoir storage products [Crétaux et al., 2011; Gao et al., 2012] is inadequate for supporting flood monitoring and forecasting purposes. On one hand, intense (i.e. flood causing) rainfall events usually occur within a short period of time (less than a week). On the other hand, depending on the satellite orbit, altimetry data are only collected every 10-35 days. Furthermore, VIS/NIR based area estimations are infeasible for flood monitoring, as floods are often associated with severe cloud contamination. An algorithm by Schwatke et al., [2015] leveraged multiple radar altimetry data sources to maximize the temporal resolution of selected reservoirs. Still, such an approach is only pertinent to a few large reservoirs which have observations from multiple radar altimeters.

In summary, increasing the spatial representativeness and temporal resolution of the remotely sensed reservoir storage variations are the two biggest challenges towards advancing the flood monitoring and prediction skills over the transboundary river basins in South Asia. Therefore, the objective of this dissertation is to generate remotely sensed reservoir storage datasets over the South Asia region with both high temporal resolution and large spatial coverage. To achieve this objective, new algorithms are needed to address the following three questions:

(1) How can we leverage the high spatial resolution satellite laser altimetry data—despite its short lifetime and long return period—for monitoring reservoir storage over the long term?

(2) How can we increase the temporal resolution of remotely sensed storage estimations under all-weather conditions?

(3) How can we stretch the spatial coverage of the reservoir storage product without being limited by the availability of satellite altimetry data?

In order to answer the above questions, the structure of this dissertation progresses from the generation of a first South Asian reservoir storage dataset (Chapter II), to enhancing the temporal resolution of the dataset (Chapter III), and then to increasing the spatial coverage of the remotely sensed reservoirs (Chapter IV).

In Chapter II, the first question was answered by developing an algorithm which successfully avoided the limitations of traditional radar altimetry sensing. Instead, the approach used water surface area estimations from the Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices product and the area-elevation relationship to estimate reservoir storage. The surface elevation measurements were from the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud and land Elevation Satellite (ICESat). In order to improve the accuracy of water surface area estimations for relatively small reservoirs, a classification enhancement algorithm was developed. This remotely sensed product contains time series data (from 2000 to 2012) of reservoir elevation, area, and storage for a total of 21 reservoirs, which represents 28% of the integrated reservoir capacity in South Asia. The product also was validated over five reservoirs where in situ observations are available.

In Chapter III, the second question was answered by leveraging the cloud penetration capability of passive microwave sensors to overcome the cloud

contamination problem associated with optical sensing. Specifically, reservoir surface area time series under all-weather conditions were derived by fusing passive microwave and optical observations. The passive microwave data are the daily horizontal brightness temperatures at 36.5 GHz and 0.25° resolution from the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E on NASA's Aqua satellite), while the optical observations are from MODIS. By calibrating against area estimations from MODIS (2003 to 2007), a set of weighting parameters were identified for each AMSR-E grid cell over the reservoir and in its vicinity. These parameters were then applied to the AMSR-E data to calculate surface area from 2008 to 2010. Storage variations are then inferred from the area and a pre-determined area-elevation relationship. Furthermore, a Kalman filtering method was performed to reduce the noise in the 4 day AMSR-E based reservoir storage dataset.

In Chapter IV, the third question was answered by extracting the elevation-area relationship from the Digital Elevation Model (DEM) data collected by the Shuttle Radar Topography Mission (SRTM), which is globally available. The approach contains three steps. First, the largest MODIS area (for each reservoir) obtained from 2000 to 2015 was overlaid onto the SRTM DEM map to delineate the region in which the water area-elevation relationship is derived. Then, the area-elevation relationship is obtained by regressing the cumulative area against the elevation values within the delineated reservoir region. Finally, the storage change values are calculated over the entire study period by applying the area-elevation relationship to the area estimations. Using the SRTM based method, the spatial coverage of the South Asian reservoir dataset can be

extended significantly. A total of 27 reservoirs are observable using this method, which represent 45% of the overall storage capacity in South Asia.

CHAPTER II

MONITORING RESERVOIR STORAGE IN SOUTH ASIA FROM MULTISATELLITE REMOTE SENSING*

2.1. Introduction

More than 45% of the global land area (excluding Greenland and Antarctica) is covered by a total of 261 international river basins [Wolf et al., 1999]. The hydrologic and political effects of transboundary rivers are enormous. Across the world, South Asia is the region that suffers the most from these impacts. Throughout much of history, certain transboundary rivers such as the Ganges, Indus and Brahmaputra rivers have served as the cultural and economic backbone of South Asia, which contains one of the largest and densest populations in the world. However, these transboundary rivers, along with the reservoirs on the rivers, are a near constant source of conflict between countries in the region. Due to these social, economic, and politically induced conflicts, countries in this region have largely failed to reach any agreements on sharing the waters of these transboundary rivers [Biancamaria et al., 2011]. Consequently, the lack of communication, particularly about reservoir storage and management, exacerbates the casualties and economic losses from flood events. Statistics based on past records show that South Asia experiences one of the highest fatality rates in the world due to floods [Adhikari et al., 2010]. Therefore, there is a strong societal need to advance the scientific

*Reprinted with permission from "Monitoring reservoir storage in South Asia from multisatellite remote sensing" by Zhang, S., H. Gao and B. S. Naz, 2014. *Water Resources Research*, doi: 10.1002/2014WR015829, Copyright 2014 by John Wiley and Sons.

understanding of the flood regimes in South Asia and provide decision makers with the information needed to better manage the reservoirs.

Satellite remote sensing has offered a unique opportunity to study the Earth from space [Rodrigues et al., 2012]. Its global coverage (which is free of geographical and political limitations) has shed light on flood monitoring and forecasting in these international river basins. One of the applications is to force hydrological models with satellite precipitation products (e.g. the Tropical Rainfall Measurement Mission (TRMM) product) to estimate river discharges [Huffman et al., 2007]. However, because reservoirs have not been implemented in hydrological models in an operational fashion at large scale, a direct result is a significant number of false alarms in basins with large reservoirs [Wu et al., 2012]. Therefore, near realtime observation data showing water storage for the reservoirs over these international basins is essential for improving hydrological modeling forecast skills and mitigating flood costs effectively.

The common approach for monitoring reservoir storage using remote sensing data is to retrieve water surface area and elevation separately, and then combine these two pieces of information for calculating the storage [Crétaux et al., 2011; Gao et al., 2012]. For measuring surface water extent, the most commonly used spaceborne instruments are the Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+) onboard Landsat, and the MODIS onboard Earth Observation System (EOS) Terra and Aqua satellites. Landsat has a high spatial resolution of 30 m with a repeat cycle of 16 days. Despite the benefits of its high spatial resolution, its low temporal resolution hampers its monitoring capability (especially when the images are impacted by cloud contamination,

which is fairly often). Compared to Landsat, MODIS has a much higher temporal resolution (i.e., daily). The tradeoff is that the spatial resolution of the MODIS sensors (250 m/500 m/1000 m) is much coarser than those of Landsat (30m). Nevertheless, the 16-day composite of MODIS images has a much larger cloud-free coverage area than that of the (once every 16 days) Landsat image. This is critical for monitoring purposes.

For estimating water surface area, index-threshold based approaches and image classification based approaches are most commonly used. Index-threshold based approaches have the advantage of estimating the water surface area with little computation. For example, the Normalized Difference Vegetation Index (NDVI) value of a water pixel is either negative or close to zero. Therefore, a threshold (e.g., zero) can be set to differentiate the water and non-water pixels on remote sensing images [Islam et al., 2010]. Some other indices, such as the modified normalized difference water index (MNDWI) [Xu, 2006], have also been used for extracting water coverage. However, because both atmospheric absorption and water quality vary by lake locations, more accurate extraction of water bodies requires the threshold to be manually adjusted according to different situations [Ji et al., 2009]. When multiple reservoirs are studied simultaneously on a large scale, they may have distinct properties and be located in different climate zones—in which case the single threshold method cannot be performed consistently for estimating the water surface area. To avoid this problem, unsupervised image classification algorithms, which do not require manual adjustment of parameters, are preferred for extracting water surface area [Song et al., 2013]. Despite their advantages when estimating water surface area at a large scale, unsupervised

algorithms—which work by minimizing the intra-cluster (i.e. within a given cluster) sum of squares—sometimes can be trapped in local optima, and the accuracy of classification results can be compromised [Maulik and Saha, 2010]. In order to improve the accuracy for water surface estimation, Gao et al. [2012] developed a new method which combines the index and classification algorithms. A classification mask was derived based on the percentile image of a set of crude classifications (using 0.1 as the NDVI threshold), and the K-means classification algorithm was applied to each NDVI image within the masked area.

Besides the surface area, water surface elevation is also needed for estimating reservoir storage. Satellite radar altimetry has been the most commonly used data source for estimating surface elevations of water bodies [Calmant et al., 2008]. There are several databases which provide elevation data from ENVISAT, GFO, Jason-1, Jason-2, and Topex/Poseidon (T/P) satellites [Birkett and Beckley, 2010; Crétaux et al., 2011; McKellip et al., 2004]. However, due to the coarse cross-track spacing (several hundred kilometers) and the relatively stretched along-track path length required to obtain accurate vertical measurements (typically 10 km or so), such elevation data are only available for a couple hundred large lakes and reservoirs. In addition to the radar altimeters, the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud and land Elevation Satellite (ICESat) has been used for measuring lake/reservoir elevations. Compared to radar altimeters, ICESat/GLAS has a high horizontal spatial resolution (approximately 70 m) and a high vertical precision ~10cm [Zhang et al., 2011]. These advantages allow ICESat/GLAS to detect elevations for much smaller water bodies (with

higher accuracy) than a typical radar altimeter can. However, the use of ICESat/GLAS for monitoring water elevations operationally has been hindered by its short life time and long repeat period (91days). During its lifetime from 2003 to 2010, the ICESat/GLAS instrument only collected elevations during designated campaigns [Wang et al., 2013]. As a result, most of the lake studies using the ICESat/GLAS focus on interannual elevation variations [Li et al., 2011; Phan et al., 2012; Shuman et al., 2006; Wang et al., 2013; Zhang et al., 2011].

Although a number of studies have used satellite data to estimate reservoir storage, the existing algorithms are all limited in different ways. For instance, in methods based on Landsat images and ICESat/GLAS data [Duan and Bastiaanssen, 2013; Song et al., 2013], the empirical relationship between the ICESat/GLAS elevation and the Landsat water surface area was established so that storage can be estimated by area when the ICESat/GLAS data was unavailable. However, the number of days when reservoir volume can be estimated is still very low because the Landsat repeats every 16 days, and its images often suffer from cloud contamination. Most of the more frequent storage observations rely on elevations from radar altimeters. Crétaux et al., [2011] had combined such elevations with area estimations from various sources (e.g., Landsat, ICESat/GLAS, Jason 1/2, T/P, ENVISAT). Due to the variety of sensor spatial/temporal resolutions and frequencies, there is a lack of consistency within the product and the product uncertainties are hard to quantify. In contrast, Gao et al. [2012] developed a global large reservoirs storage estimation algorithm which only relies on the MODIS NDVI product for estimating surface areas. The tradeoff of using the medium resolution

MODIS data (250m) is that fractional water coverage for the MODIS pixels at the reservoir borders introduces error during the unsupervised classification, which especially lowers the accuracy for reservoirs with complicated shorelines. In summary, simultaneously optimizing both the spatial/temporal resolution and the coverage remains the biggest challenge towards monitoring more reservoirs with high accuracy.

In South Asia, despite the significant benefits that would result from having near realtime reservoir storage information, this has been difficult to obtain because remotely sensed surface elevation values from radar altimeters are only available for a few large reservoirs sporadically. Although water surface elevation through ICESat/GLAS is an alternative, to our knowledge there have been no storage estimations available over South Asia based on this data source. The overarching goal of this paper is to generate a remotely sensed reservoir storage dataset in the South Asia region and validate the results with gauge observations. For this purpose, the Gao et al. [2012] MODIS area algorithm was improved such that high quality water storage estimations can be achieved using ICESat/GLAS elevation and MODIS surface area. In addition to the data analysis and results validation, storage estimation uncertainties (due to reservoir surface area retrieval algorithm parameterization and elevation measurement errors) were also quantified.

2.2. Data sources and reservoir selection

This section describes the data sources and the criteria used for selecting the reservoirs monitored in this study.

2.2.1. Satellite data

2.2.1.1. MODIS data

The MODIS data, which were used for estimating reservoir area, are available from the NASA Land Processes Active Archive Center (http://lpdaac.usgs.gov/). Specifically, we acquired the NDVI 16-day product at 250-meter spatial resolution as a gridded level-3 product in Sinusoidal projection (MOD13Q1). The reasons for choosing MOD13Q1 are: 1) its 250 m spatial resolution is highest among the MODIS products, and 2) it is less impacted by cloud contamination (when compared with the 8-day MODIS product).

2.2.1.2. ICESat/GLAS

The ICESat mission was launched in January 2003 and it ended in February 2010 [Shuman et al., 2006; Zhang et al., 2011]. The GLAS on ICESat provides global measurements of polar ice sheet mass balance, cloud and aerosol heights, land topography, and vegetation characteristics with unprecedented accuracy [Schutz et al., 2005; Zhang et al., 2011; Zwally et al., 2002]. The ICESat/GLAS Release-33 elevation data available from 2003 to 2010 were obtained through the U.S. National Snow and Ice Data Center (NSIDC, http://nsidc.org/data/icesat/). During the two mission validation periods (02/20/2003 to 03/21/2003 and 09/25/2003 to 10/04/2003), the ICESat/GLAS repeat time was eight days. However, during the rest of the mission its repeat time was 91 days. A given location along the tracks of ICESat/GLAS was typically observed less

than 24 times during the entire mission period with a vertical precision of better than 10 cm [Zhang et al., 2011; Zwally et al., 2008].

2.2.1.3. Landsat 7 Enhanced Thematic Mapper Plus (ETM+)

To evaluate MODIS based water area estimations, Landsat 7 ETM+ Level 1 data were employed to create high quality water classification images. For each of the five selected reservoirs (Section 2.2.3), a Landsat 7 ETM+ image with little cloud contamination (less than 10%) was acquired from the USGS website (http://earthexplorer.usgs.gov/). The Support Vector Machine (SVM) supervised classification approach was then used to generate the corresponding Landsat water classification image from bands 1-5, and 7.

2.2.2. Gauge observations

For validating the remotely sensed reservoir elevation and storage estimations, gauge observations reported by the Indian Operation Performance Monitoring Division of Central Electricity were employed. The online dataset (which is available at http://www.cea.nic.in/hyd_arch.html) contains daily reservoir elevation, storage, and cumulative energy generation data for 30 hydropower reservoirs from 2008 to present, with a lag time of about 2-4 months.

2.2.3. Reservoir selection

Reservoir information provided by the Global Reservoir and Dam (GRanD) database [Lehner et al., 2011] was utilized to help identify the reservoirs selected in this study. Considering the medium resolution of MODIS NDVI and the narrow ICESat/GLAS tracks, the GRanD reservoirs which meet the following criteria were

selected: 1) the area at capacity is larger than 65 km^2 which ensures that it (the surface area) can be estimated from the medium resolution MODIS NDVI (250 m) accurately; and 2) there were at least five along-track ICESat/GLAS water surface elevation measurements over the selected reservoir for each overpass (such that an average elevation—which represents the entire reservoir—can be paired up with the MODIS area).

Following the above criteria, a total of 21 reservoirs were chosen for this study, which represent 28% of the integrated reservoir capacity in South Asia (according to the GRanD database). The locations of these reservoirs are shown in Figure 2.1.

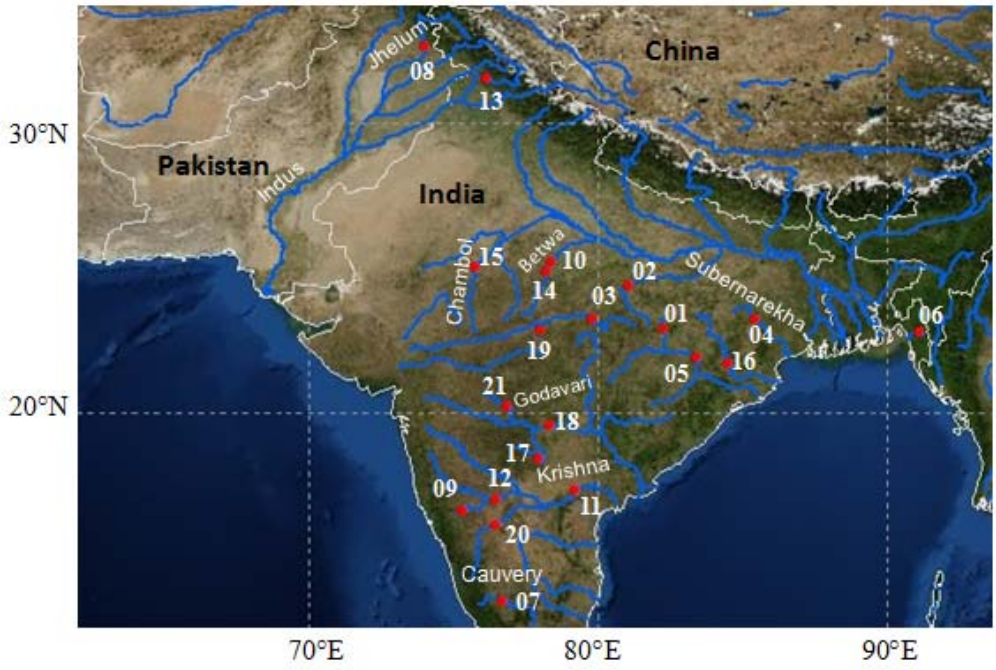


Figure 2.1 Locations of 21 selected reservoirs in South Asia. For each reservoir I.D., detailed information (e.g., name, location, capacity) is provided in Table 2.1.

2.3. Methodology

The algorithm for estimating the reservoir storage contains the following steps: 1) estimation of the water surface area of reservoirs based on MODIS NDVI images from 2000 to 2012; 2) extraction of the reservoir surface elevations from ICESat/GLAS data; 3) establishment of the area-elevation relationship for the reservoir of interest, and then retrieving the water surface elevation value from the water surface area value using the established relationship; and 4) calculation of the reservoir storage over time from water surface elevation and area time series. Figure 2.2 shows the flowchart of the algorithm. More details of the algorithm are described in the following paragraphs.

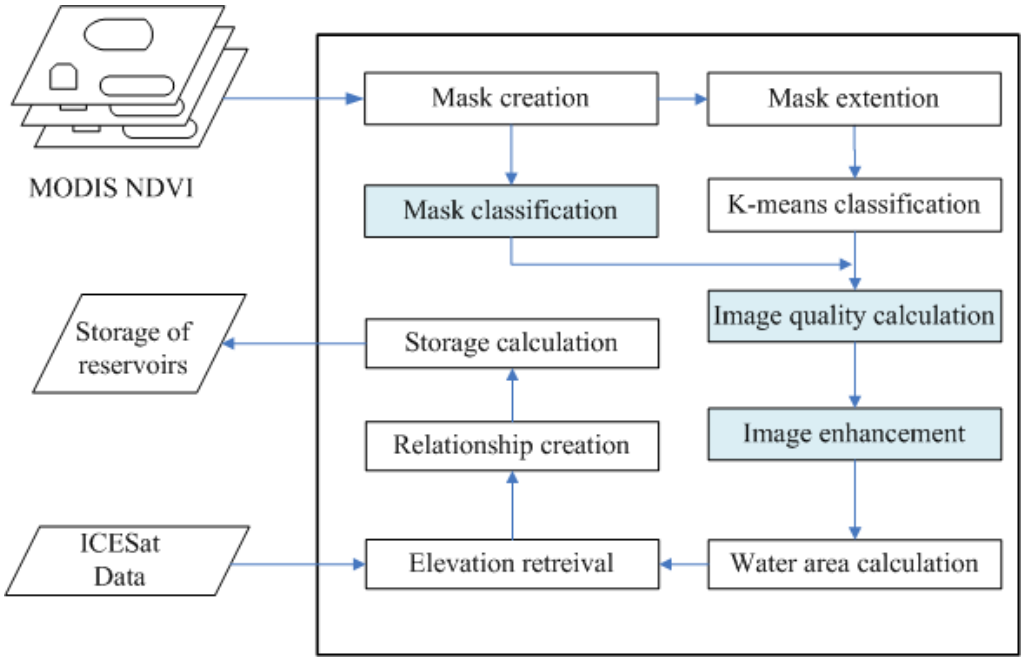


Figure 2.2 Flow chart of the reservoir storage estimation algorithm, with the image enhancement process highlighted in blue.

2.3.1. Surface area estimation

The surface area estimation method is a modification of the algorithm created by Gao et al. [2012]. In Gao et al. [2012], the storage estimations based on MODIS surface areas (using the area-elevation relationship) were not as good as the storage estimations based on radar altimeter elevations. Given the goal of this study is to use MODIS surface area (considering the limited availability of ICESat/GLAS elevation observations) as the primary input for estimating storage, it was essential to improve upon the earlier [Gao et al., 2012] MODIS area algorithm. In doing so, a few steps (as highlighted in Figure 2.2) were implemented to enhance the area retrieval accuracy and to facilitate successful retrieval—even in the case when a large portion of an image was of low data quality (e.g., cloud contaminations). Here, we explain the updated algorithm entirely with a focus on the classification enhancement procedures.

2.3.1.1. Preliminary water surface classification

First, for each 16-day NDVI image from 2000 to 2012 (296 images in total), a simplified classification was made by applying a threshold of 0.1 to each pixel—pixels with NDVI values less than 0.1 were considered as water. Based on these 296 classifications, a mask image, which represents the water coverage percentile, was created. Next, a mask extension was generated by expanding it (the mask) to a buffer area which covered any "non-water" areas that fell within a (3×3 pixels) moving window centered on each water pixel. This way the mask and its buffer area would be able to cover all possible water pixels for the studied reservoir. Figure 2.3 shows an example of the extended mask (with its buffer area included) over the Hirakud reservoir.

Finally, the K-means clustering algorithm was applied to classify all pixels of the MODIS NDVI image within the extended mask area. The pixels were divided into three classes: “water”, “non-water dry surface” and “non-water wet surface”. The classification within the mask area alone can effectively reduce the amount of computation and increase the accuracy of MODIS water area estimation. However, the accuracy of this classification approach also depends on the quality of the MODIS NDVI data. If the reliability of a pixel was not denoted as “Good Data” (i.e., the pixel was identified as “cloudy”, or “snow/ice”, or “marginal”), then it was designated as “unclassified” (i.e., neither “water” nor “non-water”). Gao et al. [2012] used a majority filter as a post classification processing mechanism to eliminate the “salt and pepper” effect. Nonetheless, this filtering does not improve the accuracy much when the low quality MODIS NDVI pixels cover a large portion of the image.

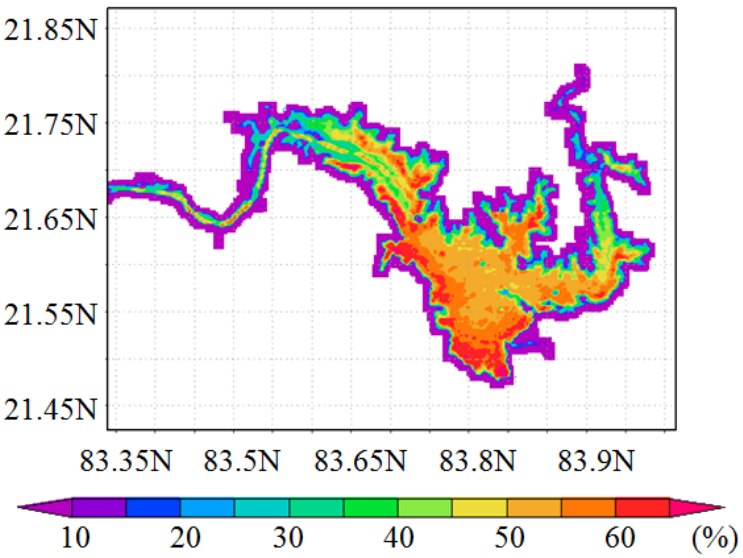


Figure 2.3 Mask of the Hirakud reservoir (with percentile for water class).

2.3.1.2. Image classification enhancement

To overcome the above limitation, a classification enhancement approach was developed in this study. The main idea is to use the percentile information from the mask image as a reference to correct misclassified pixels and to assign an appropriate class to the unclassified pixels. The steps are as follows:

First, the mask image was grouped into 50 zones based on the percentile information, using a fixed increment percentile value of 2%. This threshold (of 2%) allowed us to narrow down the differences among pixels within a given zone. In other words, all the pixels within the same zone indicate that they have a similar possibility of being covered by water. Here we use a simple synthetic example (with only three zones) to explain the concept and process. In the example, the mask image contains 8×8 pixels with the percentile values as shown in Figure 2.4. Using a threshold of 1/3 (which is one divided by the number of zones) the mask area can be divided into three different zones. Each pixel in the mask image is then assigned to a zone (zone 1, 2, or 3) according to its percentile value. For instance, since the two pixels in the top row of Figure 2.4 have the percentiles of 0.1 and 0.2 (which are both between 0 and 1/3), they are each assigned to zone 1.

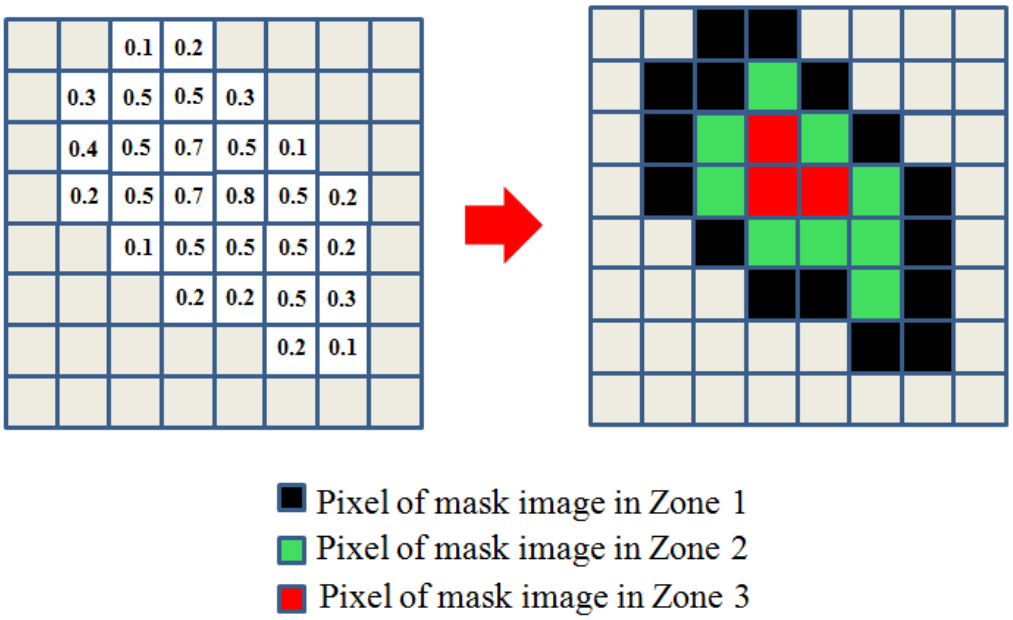


Figure 2.4 A simple example of dividing the mask into different zones according to its percentile values (Zone 1: 0~0.33, Zone 2: 0.33~0.66, and Zone 3: 0.66~1).

Second, each classification image is also divided into different zones (zone 1, 2, and 3) according to the zones of the mask image. The percentage of pixels which are classified as water within each zone (shown in Figure 2.5, rows a and b) is calculated using Equation (2.1):

$$p_i = \frac{n_i}{N_i}, \quad i = 1, 2, ..., K \tag{2.1}$$

where n_i is the number of pixles in the i_{th} zone that are classified as water (according to the MODIS NDVI classification), N_i is the total number of pixels in the i_{th} zone (according to the delineation of the mask image), and K is the total number of zones. In the simplification example, the p_i value for zone 1, zone 2, and zone 3 are 15/16, 6/9, and 0/3, respectively.

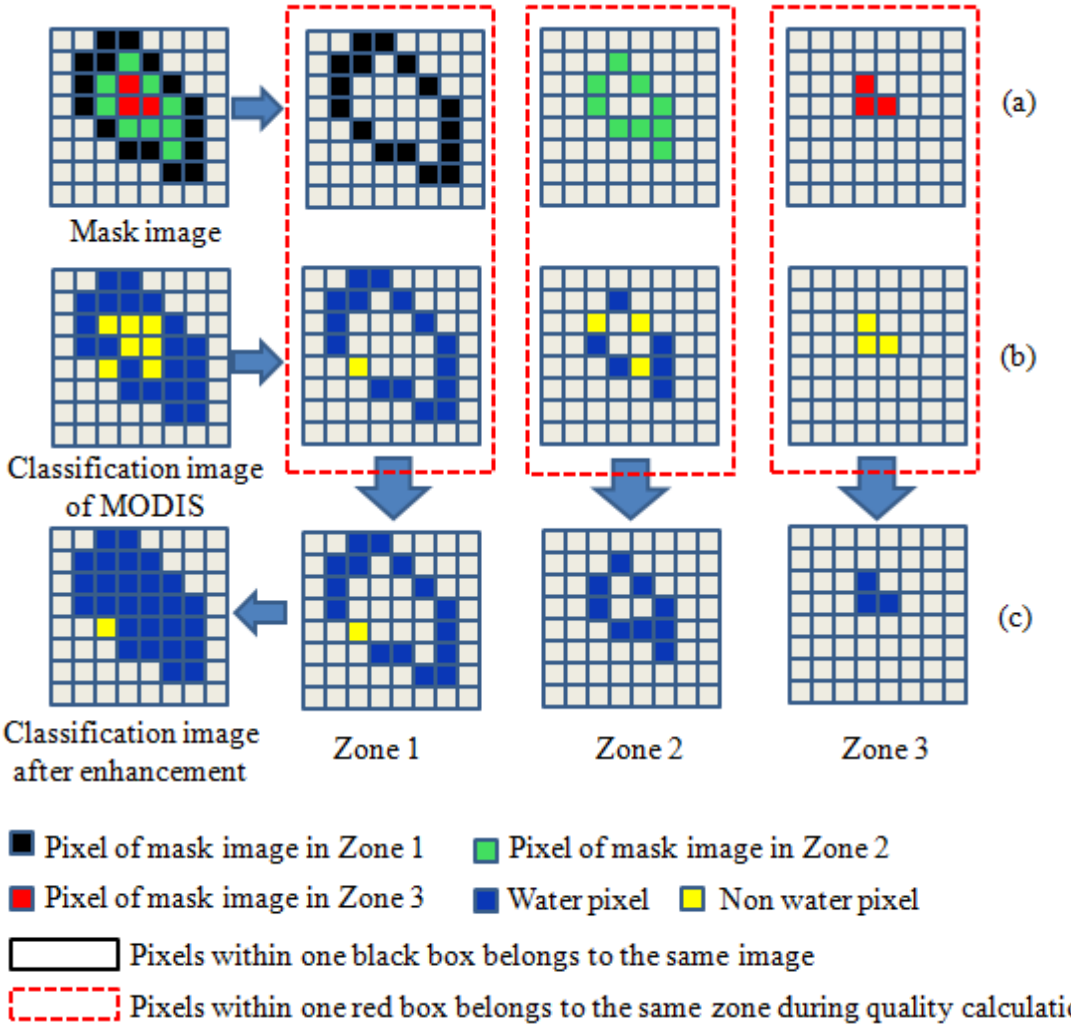


Figure 2.5 A simple example showing the classification image enhancement process: (a) dividing the mask file into multiple zones (i.e., three zones in this example); (b) assigning zone values to the classified image; and (c) enhancing the classified image based on image quality.

Third, a quality parameter (Q) is computed for each classification image according to Equation (2.2).

$$Q = \frac{\sum_{i=1}^{K} (p_i - 0.5)^2}{K} \qquad (2.2)$$

Q is a measure of the overall consistency of the surface water classification from a MODIS NDVI image. Given p_i is from 0 to 1, Q has a range between 0 to 0.25. The Q value increases as the quality of a water classification image increases. If a classification is of high quality, then the p_i values for the zones classified as water should be 1 (or close to 1), while the p_i values for the zones classified as land should be zero (or close to zero). In the case of an ideal classification (i.e., p_i equal or close to 1 or 0 for all i values), the Q value is close to the maximum (0.25). In contrast, if a classification image is of very low quality, the p_i values for most of the zones should be close to 0.5 (0.5 represents the case of a random distribution of water pixels within a zone). As a result, the Q value is close to zero for a low quality image. For the sample classification illustrated in Figure 2.5, its Q value equals to 0.156.

Last, classification image enhancement is conducted. For each zone within one classificaiton image, if its p_i value is larger than a threshod T, then all pixels in the j_{th} zone (j ranges from i+1 to K) are set as water. The threshold T is determined according to Equation (2.3):

$$T=\begin{cases} C_p & Q > C_Q \\ p_m & Q \le C_Q \end{cases} \tag{2.3}$$

where p_m is the median of all the p_i values within one classification image, and C_p and C_Q are each constant parameters. The threshold value T for each image is based on its quality Q: if Q is larger than C_Q, then T is equal to C_p; otherwise, T equals p_m. Calibrated over two reservoirs where observations are available (i.e., the Pong and Hirakud reservoirs), C_p and C_Q are set to 0.7 and 0.1, respectively. The uncertainties

associated with these parameter selections are discussed further in Section 2.4.3.1. The enhancement process for this simplified example is illustrated in Figure 2.5 (row c). Given that the Q value (Q = 0.156) of the classification image is larger than C_Q (C_Q=0.1), the threshold T is set to 0.7. For this classification image, since p_1 (p_1 = 0.94) is larger than T (T = 0.7), all pixels in zone 2 and zone 3 are assigned as water.

This classification image enhancement is based on two principles. First, a good classification image should have good consistency—meaning pixels in the same zone should have the same classification results. Second, pixels in the zones with a higher percentile should have a greater possibility of being classified as water than those in zones with lower percentile values. This means that when most of the pixels in the lower zones (usually close to the edge area) have been classified as water, pixels in higher zones (usually close to the central area) should also be water. As an example, the water classification results for the Hirakud reservoir (day 305 of 2011) are shown in Figure 2.6. The quality parameter Q for Figure 2.6b is 0.080, and the threshold T is equal to p_m (p_m = 0.64). By comparing the T value (i.e., 0.64) with the p_i values by zone, the initial classification image (i.e., Figure 2.6b) was enhanced by assigning all pixels from zone 15 to zone 50 as water (as shown in Figure 2.6c). The classification improvement due to the enhancement operation can be detected by visually comparing Figure 2.6b and Figure 2.6c.

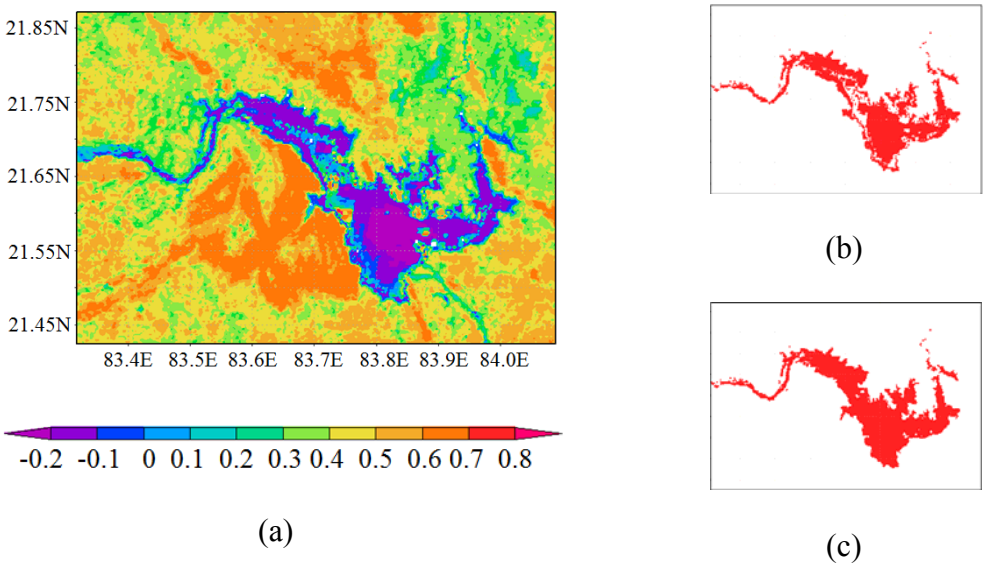


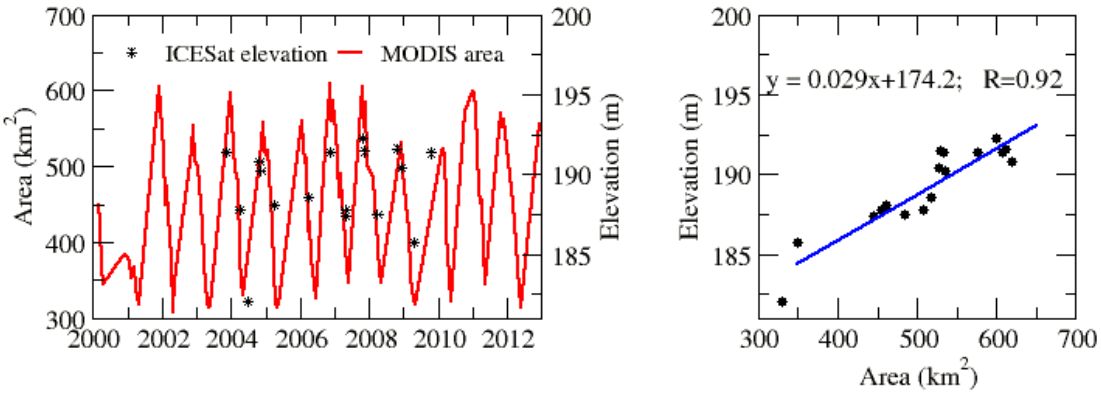
Figure 2.6 An example of the Hirakud reservoir showing the MODIS NDVI classification of day 305 of 2011: (a) the original MODIS NDVI image; (b) the classification results without image enhancement; and (c) the classification results after image enhancement.

After the preliminary classification and image enhancement are completed, the reservoir surface area can be estimated by summing up all the water pixels within the classification image (as the area for each MOD13Q1 pixel is a constant 0.25×0.25 km^2). As an example, the water surface area time series for the Hirakud reservoir is shown in Figure 2.7a.

2.3.2. Water surface elevation estimation from ICESat/GLAS

The reservoir surface elevation results were retrieved from ICESat/GLAS orbital measurements in two steps. First, using the reservoir area boundary identified by the MODIS water classification (that was closest in time with the ICESat/GLAS overpass) and the ICESat/GLAS orbital geographical location information, all elevation measurements within the reservoir were extracted. Then, the representative elevation of

a reservoir for a given day was estimated as the average of all the measurements within the overpassing orbit.



(a) (b)

Figure 2.7 The area-elevation relationship over the Hirakud reservoir: (a) time series of MODIS surface water area and ICESat/GLAS surface elevation; and (b) Scatter plot for the area-elevation relationship and the linear regression result.

2.3.3. Area-elevation relationship

The water area and elevation data collected during the overlapping period were paired up for each reservoir to get reservoir specific area-elevation relationships. Figure 2.7b shows one example of the area-elevation relationship for the Hirakud reservoir. By creating area-elevation relationships, we could use the MODIS based water surface area to estimate the water surface elevation when the ICESat/GLAS data were not available. Table 2.1 shows the area-elevation relationships and correlation coefficients for all the reservoirs in this study.

Table 2.1 Reservoir Area-Elevation Relationships and Correlation Coefficients for the 21 Reservoirs

I.D.	Reservoir	Location	Capacity (mcm)	Purpose[a]	Area-elevation relationship[b]	R
01	Bango	22.61,82.60	3416	I,E	y=0.238x+332.0	0.84
02	Bansagar	24.19,81.29	5410	I,E	y=0.051x+318.6	0.99
03	Bargi	22.95,79.93	3920	I,E	y=0.101x+398.9	0.98
04	Chandil	22.98,86.02	1961	I,E	y=0.14x+171.4	0.88
05	Hirakud	21.52,83.85	8100	I,E	y=0.029x+174.2	0.92
06	Karnafuli	22.5,92.23	6477	I,E,F	y=0.073x+13.67	0.90
07	KrisharajaSagar	12.42,76.57	1369	I,E,W	y=0.277x+729.2	0.94
08	Mangla	33.13,73.64	7300	I,E,F	y=0.243x+311.7	0.80
09	Malaprabha	15.82,75.09	1068	I,E	y=0.112x+612.6	0.98
10	Matatila	25.10,78.37	1133	I,E	y=0.112x+297.7	0.95
11	Nagarjuna Sagar	16.57,79.31	11600	I,E	y=0.379x+100.2	0.95
12	Narayanapura	16.22,76.35	1071	I	y=0.140x+478.6	0.88
13	Pong	31.97,75.95	8570	I,E	$y=0.237x^2$-188.6x +37675	0.98
14	Rajghat	24.76,78.23	2172	I,E	y=0.085x+352.0	0.99
15	RanaPratapSagar	24.92,75.58	2898	I,E	y=0.134x+325.5	0.97
16	Rengali	21.28,85.03	3168	I	y=0.072x+100.2	0.94
17	Singur	17.75,77.93	850	W	y=0.082x+513.1	0.97
18	SriramSagar	18.97,78.34	3172	I,E	y=0.038x+320.4	0.99
19	Tawa	22.56,77.98	2310	I	y=0.142x+335.6	0.99
20	Tungabhadra	15.27,76.33	3764	I,E	y=0.05x+481.1	0.87
21	Yeldari	19.72,76.73	934.3	I,E	y=0.459x+41.6	0.99

[a] I is irrigation, E is electricity generation, W is water supply, and F is flood control

[b] y is elevation and x is area

2.3.4. Water storage estimation

Since the ICESat/GLAS elevation data were very limited, we could not combine the ICESat/GLAS water elevation and the MODIS water surface area directly to calculate the water storage. Instead, the elevation was inferred from the MODIS surface area and the area-elevation relationship. The storage was then estimated using Equation (2.4):

$$V_{RS}=V_c-(A_c+A_{RS})(h_c-h_{RS})/2 \qquad (2.4)$$

where V_c, A_c, and h_c represent storage, area, and water elevation at capacity, and V_{RS}, A_{RS}, and h_{RS} are the estimated storage, area, and water elevation from remote sensing. Figure 2.8 shows the time-series of reservoir storage for the Hirakud reservoir as an example.

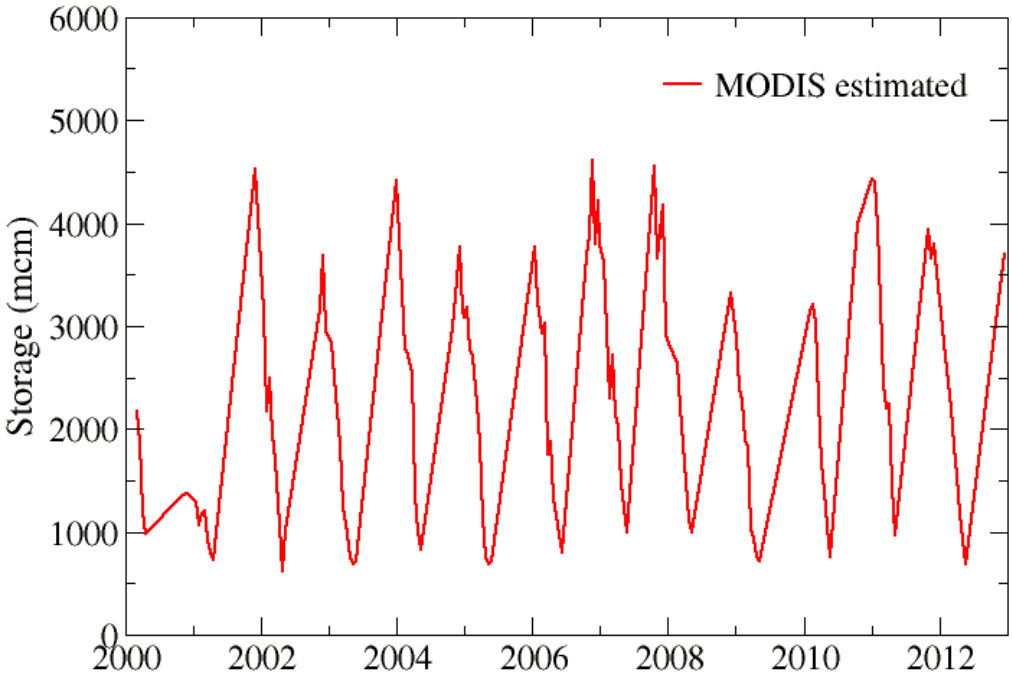


Figure 2.8 Time series of the MODIS based storage estimation for the Hirakud reservoir.

Using the methods explained in this section, time series values of reservoir surface area, surface elevation, and water storage were estimated for the 21 selected South Asian reservoirs from 2000 to 2012.

2.4. Results

2.4.1. Results validation

The remotely sensed results were validated comprehensively through two steps. First, the MODIS surface water classification images were compared with Landsat high resolution (30 m) classifications. Second, the reservoir elevation and storage dataset from remote sensing were evaluated with gauge observation. Five reservoirs (i.e., R. P. Sagar, Hirakud, Nagarjuna Sagar, Pong, and Rengali) were selected for the validation where observational data from 2008 to 2012 are available.

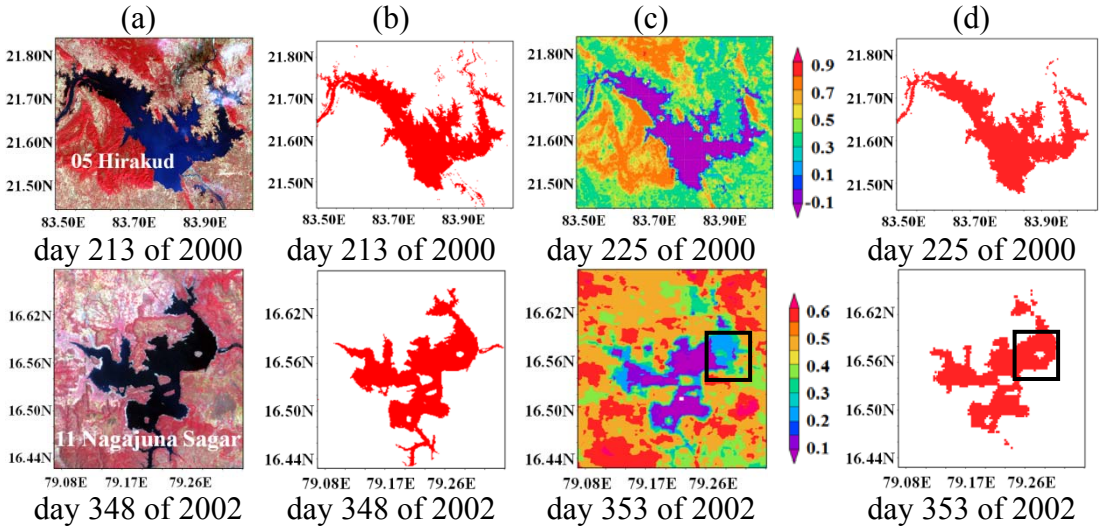


Figure 2.9 Comparisons between Landsat and MODIS surface water classification results: (a) Landsat ETM+ images (RGB); (b) Landsat classifications; (c) MODIS NDVI images; (d) MODIS classifications.

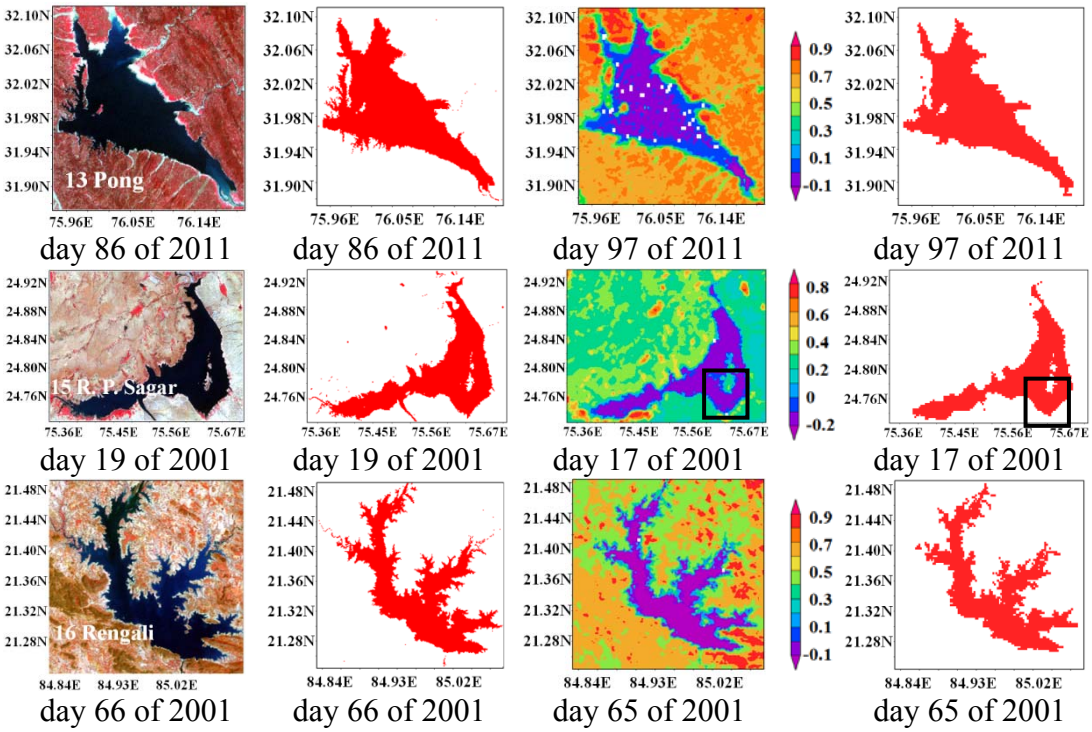


Figure 2.9 (continued).

Figure 2.9 shows the original Landsat false color images, Landsat classifications, MODIS NDVI images (which overlapped with the Landsat date), and MODIS classifications. The area estimations are summarized in Table 2.2. The MODIS classifications are in good agreement with the Landsat results, with percent error values ranging from 1% to 9%. The main differences are attributed to the different spatial resolutions. The enhanced MODIS classification algorithm also showed good performance when some pixels of the original NDVI images were of low quality (i.e., the Nagarjuna Sagar, R.P. Sagar, and Pong reservoirs). For the Nagarjuna Sagar reservoirs, pixels in the black boxes would have been misclassified as "non-water" if the enhancement operation was not performed. Noise in the Pong reservoir NDVI image

was also effectively removed in final MODIS classification image. For the R. P. Sagar reservoir, there is an island within the region marked by a box. Due to the MODIS NDVI quality issue, the island in the original MODIS image is much smaller than that in Landsat. The underestimated island area was corrected in the enhancement operation by rectifying the misclassification using the historical percentile coverage information.

Table 2.2 Comparisons between Landsat and MODIS Water Surface Area Estimations

	Hirakud	Nagarjuna Sagar	Pong	R.P.Sagar	Rengali
Landsat area (km^2)	572	185	179	160	189
MODIS area (km^2)	554	179	195	158	195
percent error[a] (%)	3	3	9	1	3

[a] percent error is defined as |MODIS area - Landsat area|/Landsat area × 100%

To evaluate the elevation and storage estimations, observation data were obtained from the Indian Century Electricity Authority (CEA), (http://www.cea.nic.in/hyd_arch.html), which provides daily observed water storage (V_o) and water level (h_o) for 30 hydropower reservoirs (with roughly a 2-4 month lag time). For comparison purposes, another remotely sensed dataset was generated over these five reservoirs from the same data source (i.e., MODIS and ICESat/GLAS) using the algorithm by Gao et al. [2012]. Figure 2.10 shows the validation results of these two algorithms. Among the five reservoirs, both algorithms performed the best over the Hirakud. This is because the Hirakud has the largest area (among the five evaluated reservoirs) and it is surrounded by heavy vegetation whose NDVI values are very distinctive from those of water. For the Pong reservoir, Gao et al. [2012] 's algorithm tends to underestimate when the storage was large while overestimate when the reservoir

was small. Examination of a series of MODIS classification images for this reservoir using both algorithms (results are not shown) suggest that the Gao et al. [2012] approach had overestimated the area when the reservoir was about half full. Since the Gao et al. [2012] approach created two classes (i.e., "water" and "non-water"), the wetland was often misclassified as "water". But when the reservoir was mostly full or mostly empty, the wetland was either very small or had a great contrast to the water. A direct consequence of this is a skewed area-elevation relationship - which ultimately led to the elevation and storage errors. To overcome this problem, we designated three classes (with "wetland" as a standalone class) instead of two. For this particular reservoir, a large wetland was created when the water level retreated to its middle range.

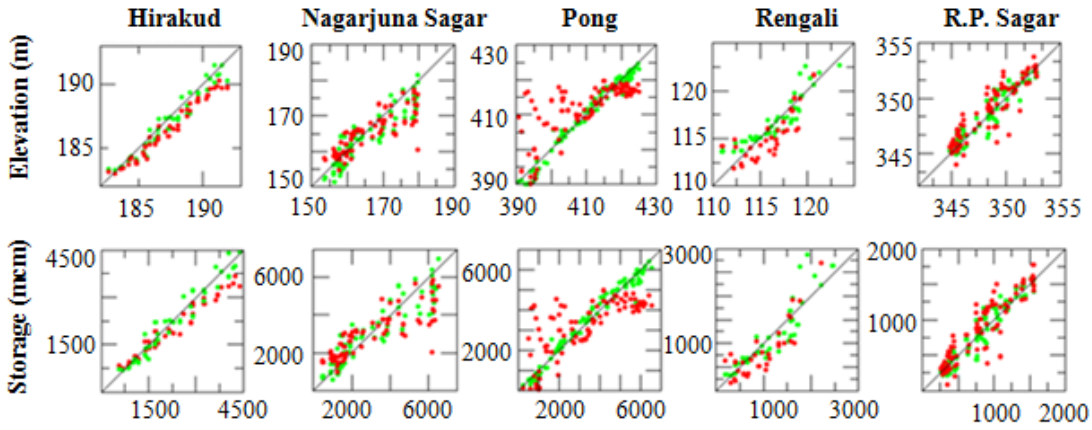


Figure 2.10 Validations of the remotely sensed water surface elevation and storage data using gauge observations over five reservoirs. The x-axis is the observation data and the y-axis is the remotely sensed result. The green dots represent results obtained from the enhanced algorithm in this paper and the red dots are results based on the Gao et al, 2012 algorithm.

In addition, three statistical criteria were examined in Table 2.3. They are the correlation coefficient (*R,* Equation (2.5)), bias (*B,* Equation (2.6)), and Normalized root-mean square error (NRMSE, Equation (2.7)).

$$R = \sqrt{\frac{Cov(RS, Obs)}{D(RS)D(Obs)}} \tag{2.5}$$

$$B = \overline{RS} - \overline{Obs} \tag{2.6}$$

$$NRMSE = \frac{\sqrt{\sum_{i=1}^{n} \frac{(RS_i - Obs_i)^2}{n}}}{\overline{Obs}} \tag{2.7}$$

where *RS* is the result from remote sensing, *Obs* is the observation data, n is the number of data points, and $\overline{RS}$ and $\overline{Obs}$ are the average values of the remote sensing result and the observational data. *Cov()* means covariance and *D()* represents the variance.

According to the statistics, the new algorithm outperformed the Gao et al. [2012] algorithm in all cases except for the elevation bias at Nagarjuna Sagar. Taking the NRMSE of storage as an example, this algorithm led to an improvement (over Gao et al., [2012]) of 6.10%, 11.77%, 28.26%, 18.40%, and 8.76% for the Hirakud, Nagarjuna Sagar, Pong, R.P. Sagar and Rengali, respectively. As explained in Sections 2.3.1.1 and 2.3.1.2, the key differences between these two algorithms are the number of classes used in the K-mean unsupervised classification and the post classification filtering procedure. The algorithm used in this study out performed Gao et al. [2012] in both accounts. First, the generation of three classes meant that the misclassifications of the pixels along the shore were avoided. Second, the post classification enhancement employed a more

realistic weighted procedure which took into account which percentile zone a pixel belonged to (vs. the simple "majority filter" method, which treats all pixels evenly across the entire classification image).

Table 2.3 Statistical Validation Results for the Remotely Sensed Reservoir Elevation (*h*) and Storage (*V*)

		Hirakud		Nagajuna Sagar		Pong		Rengali		R.P. Sagar	
		Gao[a]	Zhang[b]	Gao	Zhang	Gao	Zhang	Gao	Zhang	Gao	Zhang
R	*h*	0.98	0.98	0.83	0.92	0.74	0.99	0.79	0.91	0.90	0.96
	V	0.96	0.97	0.83	0.92	0.75	0.99	0.84	0.92	0.90	0.96
Bias	*h* (m)	-0.84	-0.34	-1.20	-1.45	2.02	0.22	-0.47	0.33	0.42	0.22
	V (mcm)	-284.50	-98.10	-136.30	-32.30	138.69	10.70	-133.90	6.08	47.47	12.35
NRMSE	*h* (%)	11.28	7.43	10.61	6.78	28.32	6.45	16.27	12.68	6.87	4.45
	V (%)	18.87	12.77	36.52	24.75	37.77	9.51	43.60	25.20	24.21	15.45

[a] Gao refers to the algorithm in Gao et al. [2012]
[b] Zhang refers to the algorithm in this paper

The results from Table 2.3 also suggest multi-criteria should be considered for a comprehensive evaluation. Although a low correlation indicates low accuracy, a high correlation does not necessarily mean there is not a problem. For instance, the *R* values from both algorithms are very high over the Hirakud, but the biases using the Gao et al. [2012] algorithm were much larger than those from our algorithm due to underestimations. Another example is the Nagarjuna Sagar reservoir. Due to error cancellation, the bias of elevation at the Nagarjuna Sagar reservoir was smaller from the Gao et al. [2012] algorithm. However, when *R* and NRMSE are both considered, the method proposed in this study performs better.

2.4.2. Reservoir storage variations in South Asia

Figure 2.11 shows the storage variations for the 21 studied reservoirs in South Asia from 2000 to 2012. The total capacity of theses reservoirs is 83,925 million cubic meters (mcm). The record length for all reservoirs is 13 years except for the Bansagar reservoir (where construction was not completed until 2006). A few examples that explore the hydrological implications of these storage time series are as follows.

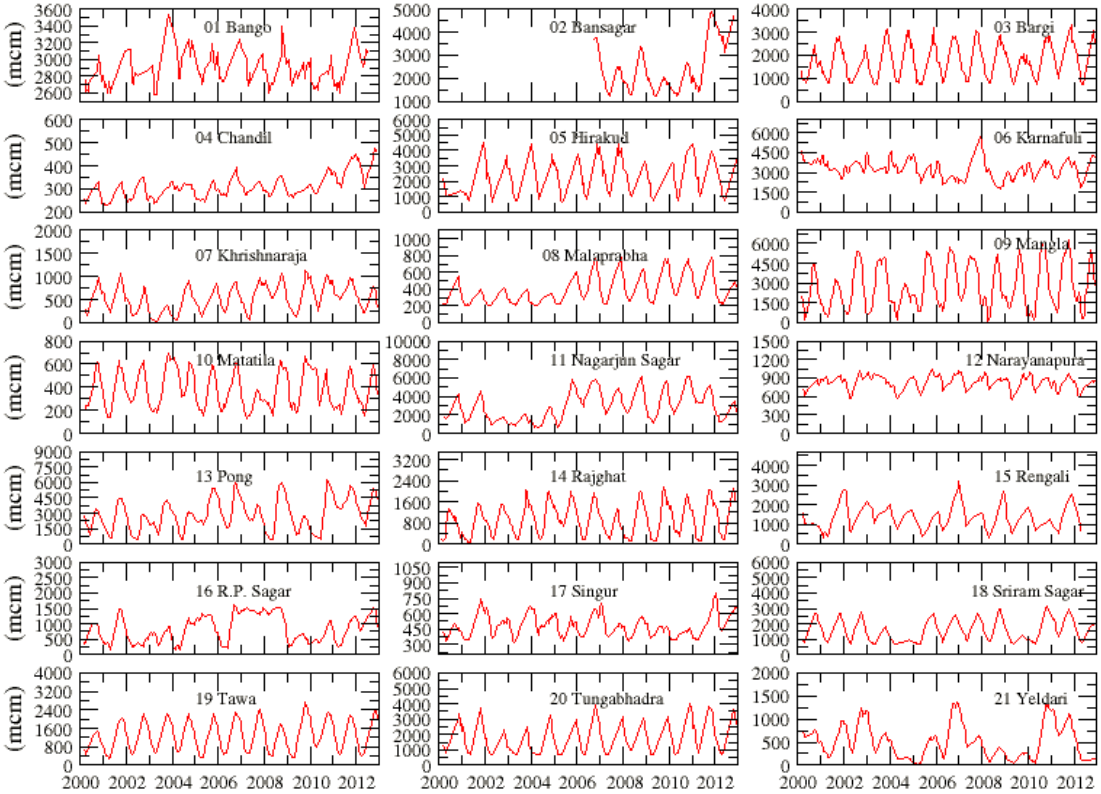


Figure 2.11 Remotely sensed storage time series for the 21 South Asian reservoirs.

Built in 1975, the Pong reservoir (i.e. Maharana Pratap Sagar) is the highest earthfill dam in India. It is located on the Beas River within the wetland zone of the Siwalik Hills (i.e.. the Outer Himalayas) in the state of Himachal Pradesh. Because it is close to Himachal Mountain, upstream snow and glacial melt contributes substantially to the lake inflows. During the 13-year study period, a storage increase trend of 133 mcm/yr (133 mcm is equivalent to 1.9% of the reservoir capacity) was identified. This trend is attributed to global warming, which increases the snow and glacial melt in the High Asia region and leads to above normal river discharge to downstream lakes [Moors and Stoffel, 2013]. Meanwhile, this region is prone to floods during the monsoon season. According to the Global Active Archive of Large Flood Events database [Brakenridge et al, 2002], major floods in August of 2001 and 2007 caused 16 and 76 fatalities, respectively. These flood events are well represented by the peaks shown in the Pong reservoir storage time series in Figure 2.11. Also, in January 2010 the storage was only 20 percent of capacity, which reflects the 2009-2010 drought in the Himachal area.

Yeldari reservoir, another earthfill dam in India, is mainly used for irrigation and hydroelectricity generation. According to media reports, two severe drought events occurred in the region in 2004 and 2012, and the Yeldari dam reservoir almost dried up in both cases ("38 reservoirs down to 30 per cent storage" from *Rediff Bussiness*, http://www.rediff.com/money/report/water/20040728.htm, 2004; and "Marathwada remains parched" from *Afternoon Despatch & Courier*, http://www.afternoondc.in/city-news/marathwada-remains-parched/article_65090, 2012). The remotely sensed storage of the Yeldari has demonstrated clear consistency with the reported results. Moreover,

the satellite estimated time series indicated that the 2004 drought lasted until the beginning of 2005, and that there was another drought with smaller magnitude in 2009.

Another example is the Bansangar reservoir, which is located on the Sone River and is used for irrigation and hydroelectricity generation. The dam construction was started in 1978 and completed in 2006. According to Figure 2.11, Bansagar reservoir water storage kept increasing in 2011 until the fall season. Following the sudden water release into the Sone river from the reservoir and the heavy rainfall in its downsream area, several villages in Rohtas, Arwal, Patna, Aurangabad and Bhojpur districts were reportedly inundated ("Flood alert sounded in Bihar" from *The Hindu*, http://www.thehindu.com/todays-paper/tp-national/tp-newdelhi/flood-alert-sounded-in-bihar/article2489034.ece, 2011, *media report*).

The upper drainage basin of the Mahanadi River is characterized by periodic droughts, which is a contrast to the lower delta region where floods are common. The Hirakud reservoir was constructed to help alleviate these problems by regulating river flows. The reservoir also produces hydroelectricity through several hydroelectric plants. In 2000, the region suffered from historical drought, which is effectively reflected by the low storage values (Figure 2.11). Although the reservoir storage is directly affected by inflows, the operation rules play an important role in regulating the storage. It was reported that Hirakud kept the water level higher than the recommended value in 2008. When inflows increased suddenly, the Hirakud released water (in order to protect the dam) which led to a man-made flood in the downstream area [Choudhury et al., 2012]. Although 2009 and 2010 were two dry years, the Hirakud reservoir peak storage

maintained about 70% of capacity. When heavy precipitation occurred in September 2011, the mismanagement of the Hirakud reservoir caused avoidable flooding [Choudhury et al., 2012].

2.4.3. Uncertainty analysis

In this section, uncertainty analysis for the remotely sensed storage estimations was conducted. The sources of storage estimation error include ICESat/GLAS elevation error, MODIS water surface area error, area-elevation relationship error, and the reservoir configuration error. Specifically, we investigated the uncertainty of storage associated with two sources: the surface area estimation and the ICESat/GLAS elevation data. This way the area-elevation relationship error is addressed implicitly. The reservoir configuration error, which is both hard to quantify and less relevant to the estimation algorithm in this study, is not discussed.

2.4.3.1. Uncertainty due to parameter selection in water area classification

During the water area classification process (as described in Section 3), parameterization uncertainty is related to the selection of two parameters: C_P and C_Q. In Equation (2.3), C_P is set to 0.7 based on "trial and error" over the Pong and Hirakud reservoirs. When Cp is set to a lower value (e.g., 0.5), a non-water pixel will have a higher possibility to be assigned as water (according to Equation (2.3)). In order to test the uncertainty of C_P, we calculated the storage difference between $C_p = 0.5$ and $C_p = 1$ for each of the 21 reservoirs.

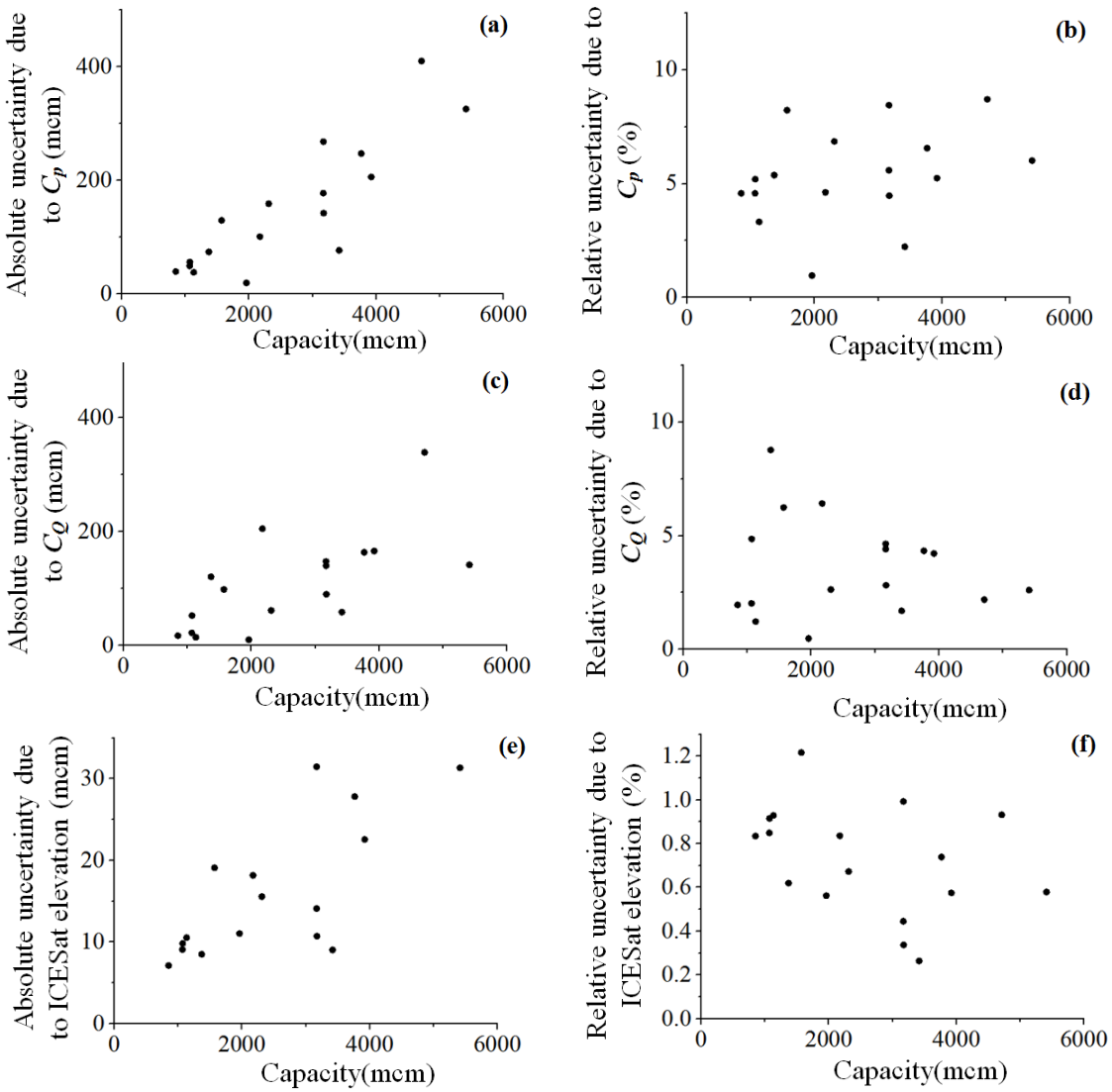


Figure 2.12 Uncertainty analysis results: (a) absolute uncertainty due to C_p in water area classification; (b) relative uncertainty due to C_P in water area classification; **(c)** absolute uncertainty due to C_Q in water area classification; (d) relative uncertainty due to C_Q in water area classification; (e) absolute uncertainty due to ICESat elevation observations; and (f) relative uncertainty due to ICESat elevation observations.

Figures 2.12a and 2.12b show the absolute uncertainty and relative uncertainty due to the choice of C_p. The absolute uncertainty increases as the reservoir capacity increases, while the relative uncertainly has no clear relationship with the capacity. The average relative uncertainty is 5.54%. The second parameter, C_Q, is used as a criterion to identify

whether a classification should be considered "good" or "poor". When C_Q is set to 0 (i.e., the minimum of Q), the final classification results (after the enhancement operation) for most of images will be similar to the results derived by the K-means algorithm, which tends to underestimate the water surface area. However, when C_Q is set to 0.25 (i.e., the maximum of Q), although noise can be easily deleted, the water surface area may be overestimated. The difference between the storage estimation values when using $C_Q = 0$ and $C_Q = 0.25$ represents the uncertainty of a given reservoir. The absolute uncertainty and relative uncertainty associated with C_Q for the 21 reservoirs are shown in Figure 2.12c and Figure 2.12d, respectively. Although the absolute uncertainty due to C_Q shares a similar pattern with that due to C_p, the relative uncertainty due to C_Q converges to a small value (of 2-3%) as the reservoir size increases. Overall, the choice of C_p adds more uncertainty than the selection of C_Q. Nonetheless, these relative uncertainties are smaller than the NRMSE values in Table 2.3.

2.4.3.2. Uncertainty due to ICESAT/GLAS elevation

Each elevation data record that was used for the area-elevation relationship represented the average of all ICESat/GLAS observations over a given reservoir. Therefore, there is an uncertainty associated with this averaged elevation. This uncertainty could be from sensor measurement errors and/or natural variations (including surface roughness and surface wind). Equation (2.8) shows the uncertainty of storage due to elevation uncertainty:

$$\Delta V = (A_c + A_{RS})\Delta h / 2 \qquad (2.8)$$

where ΔV is the uncertainty due to the ICESat/GLAS elevation, A_c is the reservoir area at capacity, and A_{RS} is the estimated area from MODIS. Δh is the difference between the maximum and minimum ICESat/GLAS elevation values over the water surface along the track on the observation day. Unlike the quantification of uncertainties due to parameterization, the calculation of uncertainty due to elevation is based on analyzing real observations from ICESat/GLAS. Because of the relatively high accuracy of ICESat/GLAS elevation data (less than 10cm according to [Zhang et al., 2011]), uncertainty of storage caused by elevation is much smaller than that caused by area estimation. Although the absolute uncertainty in Figure 2.12e has a similar trend as those shown in Figure 2.12a and 2.12c, it is about an order of magnitude smaller. For relative uncertainty, there was no noticeable trend when the size of the reservoir changed. The relative uncertainty due to ICESat/GLAS elevation ranged from 0.34% to 1.67%, and its mean value was 0.67% (Figure 2.12f).

2.5. Summary and conclusions

In this study, a remotely sensed reservoir dataset in South Asia, which includes elevation, area, and storage information, was generated using a novel multi-satellite algorithm. First, the MODIS derived water classifications and ICESat/GLAS data (when available) were used to create an area-elevation relationship for each of the 21 selected reservoirs. Next, the elevation and storage variations were estimated over the period of 2000-2012 using the MODIS based surface area time series and the area-elevation relationships. The ICESat/GLAS has a much higher spatial resolution (70 m) than the satellite radar altimetry data (several kilometers), which allows it to measure much

smaller reservoirs. However, in the past its short lifetime (2003 to 2009) and low repeat frequency (90 days) had limited the capability of ICESat/GLAS with regards to monitoring lakes and reservoirs. By combining MODIS and ICESat/GLAS for reservoir storage estimations, we were able to take the advantage of both satellites. Furthermore, this satellite-based reservoir dataset was validated by both high resolution Landsat ETM+ classifications and gauge observations over five locations. Last, we also conducted uncertainty analysis for the remotely sensed storage estimations. Specifically, we investigated the uncertainty of storage associated with two factors: the surface area classification parameterizations and the ICESat/GLAS elevation measurements. The conclusions of this study are summarized as follows:

(1) The post-classification image enhancement procedure significantly improved the MODIS water area estimation accuracy, which is essential for the area driven (vs. elevation driven) storage estimation algorithm used in this study.

(2) By using MODIS area estimations and ICESat elevations for deriving the reservoir area-elevation relationship, the retrieval algorithm developed in this study has the potential applications in other regions where reservoir storage information is hard to acquire and radar altimetry observations are few.

(3) Uncertainty analysis results suggest that the uncertainties associated with the area algorithm parameter selections are larger than those due to elevation measurements. Nonetheless, the uncertainties are less than 10% in all cases.

(4) Considering the abundance of transboundary rivers in this region, this reservoir storage dataset can serve as a valuable data source for water resources management

purposes, such as hydropower generation, irrigation water supply allocations, and disaster mitigations. By incorporating remotely sensed reservoir storage information into hydrological modeling, better model prediction skills are expected (and false alarms can be avoided).

Although this storage dataset represent 28% of the integrated reservoir capacity in South Asia—which is first of its kind to the best of our knowledge—observations over more reservoirs would be highly valuable for this region (with its dense river networks). Unfortunately, due to the relatively sparse orbital coverage of ICESat/GLAS and the relatively coarse spatial resolution of MODIS, the reservoirs were limited to what was presented by this study. With the launch of ICESat2 mission in 2016 and the Surface Water and Ocean Topography (SWOT) mission in 2021, a significantly greater number of reservoirs will be able to be studied at much higher spatial resolutions.

CHAPTER III

A NOVEL ALGORITHM FOR MONITORING RESERVOIRS UNDER ALL - WEATHER CONDITIONS AT A HIGH TEMPORAL RESOLUTION THROUGH PASSIVE MICROWAVE REMOTE SENSING*

3.1. Introduction

It has been demonstrated that effective reservoir operations can reduce flood inundation significantly [Mateo et al., 2014]. However, in many regions of the world the sharing of storage information of reservoirs is limited (especially in developing countries), which critically hinders the capability of reservoir based flood mitigation [Adhikari et al., 2010]. Over the last few decades, the advent of satellite remote sensing has offered an unprecedented promise of monitoring lakes and reservoirs from space [Alsdorf et al., 2007; Lettenmaier et al., 2015]. The common approach for estimating storage is to combine remotely sensed surface elevation/area with reservoir bathymetry, acquired either from prior *in situ* data or from remote sensing retrievals [Duan and Bastiaanssen, 2013; Gao et al., 2012; Song et al., 2013; Zhang et al., 2014]. Satellite radiometers have been employed for measuring the water levels of large lakes and reservoirs globally [Alsdorf et al., 2001; Crétaux and Birkett, 2006], and the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud and land Elevation Satellite (ICESat) has been applied to many relatively smaller natural lakes at a regional scale

*Reprinted with permission from "A novel algorithm for monitoring reservoirs under all - weather conditions at a high temporal resolution through passive microwave remote sensing" by Zhang, S. and H. Gao, 2016. *Geophysical Research Letters*, doi: 10.1002/2016GL069560, Copyright 2016 by John Wiley and Sons.

[Smith et al., 2009; Zhang et al., 2011]. Water surface area is usually estimated from visible or near infrared (VIS/NIR) imagery.

Although remotely sensed reservoir storage time series have contributed to evaluating long term hydrologic model simulations [Zhou et al., 2016], their use for improving flood monitoring and forecasting is as of yet unexplored [Wu et al., 2012]. Two key constraints have plagued the use of near realtime satellite reservoir products with regard to water management decision making. The first constraint is attributed to the limited spatial coverage of radar altimeters. Given the capabilities of past and current radar altimeters, only some of the large reservoirs—those that are both on the satellite tracks and not affected by local topography—can be monitored. A product by Gao et al. [2012] included 34 large reservoirs over 24 river basins, representing 15% of global reservoir capacity. Since most regulated rivers contain multiple dams, information from one or two large reservoirs per basin is not adequate to fully constrain most hydrological models [Zhou et al., 2016]. To characterize the storage variations of relatively small reservoirs, a study was performed over South Asia [Zhang et al., 2014] that combined MODIS area estimations with ICESat laser altimeter measurements—in lieu of satellite radar altimeter data. The mean storage at capacity of these 21 regional reservoirs is only 3.8 km^3, which is a considerable improvement in resolution over radar altimeter based reservoir estimations by Gao et al., [2012] (which have an average capacity of 34 km^3 per reservoir).

The second constraint is the low temporal resolution of the remotely sensed reservoir products. On the one hand, intense (i.e. flood causing) rainfall events usually

occur within a short period of time (less than a week). On the other hand, depending on satellite orbits, altimetry data are only collected every 10-35 days. Furthermore, VIS/NIR based area estimations are infeasible for flood monitoring, as floods are often associated with severe cloud contamination. An algorithm by Schwatke et al. [2015] leveraged multiple radar altimetry data sources to maximize the temporal resolution of selected reservoirs. Still, such an approach is only pertinent to a few large reservoirs which are covered by multiple radar altimeters over the same time period. The unresolved question is: how can we increase the temporal resolution of remotely sensed storage estimations without sacrificing the spatial coverage?

The objective of this paper is to develop a new algorithm for monitoring reservoirs at high temporal resolution under all-weather conditions. This can be attained by leveraging the passive microwave radiative properties through a data fusion approach. Benefitting from its sensitivity to fractional water coverage and its capability of penetrating through clouds [Brakenridge et al., 2007; Gao et al., 2006; Watts et al., 2012], passive microwave remote sensing has made notable contributions for mapping wetlands and flood plains [Fluet-Chouinard et al., 2015; Papa et al., 2010; Papa et al., 2006]. However, because of their inherent coarse spatial resolution, passive microwave satellite observations have never been employed for inferring the area of any individual reservoir to date [Gao, 2015].

The novelty of this new algorithm is that it utilizes passive microwave observations—both over a reservoir, and in its vicinity—for estimating the water surface area of an individual reservoir. The algorithm assumes that a set of optimal weight

coefficients can be identified using the training data, such that the ratio between the weighted brightness temperatures of the reservoir pixels and those pixels in their vicinity is linearly correlated to the reservoir area. Specifically, we demonstrate the algorithm over four South Asian reservoirs where a prior existing remotely sensed reservoir dataset [Zhang et al. 2014] is used for training the passive microwave observations—and where *in situ* observations are available for validating the results. Because the algorithm is exclusively based on satellite data, it is applicable at a global scale. The benefits, limitations, and applicability of this algorithm are also discussed.

3.2. Data and methods

3.2.1. Remote sensing and gauge observation data

To estimate the water surface area changes, AMSR-E daily brightness temperatures at 0.25º were obtained from the National Snow and Ice Data Center (NISDC) [Knowles et al., 2006]. The frequency used in this study is 36.5 GHz. Since there is no significant interference from clouds at this frequency, it has been used in previous studies to monitor land surface temperature and surface water fraction [Fily et al., 2003; Holmes et al., 2009].

A validated 16-day remotely sensed reservoir storage product in South Asia [Zhang et al., 2014] is utilized as a training dataset for estimating reservoir area from AMSR-E observations. The training dataset was derived from the MODIS vegetation indices product (MOD13Q1) and from surface elevation measurements obtained from GLAS/ICESat. It contains water surface area, elevation, and storage data for 21 reservoirs in South Asia from 2002 to 2012, and it also provides the area-elevation (*A*-*H*)

relationship for each of these reservoirs. Validation against *in situ* observations shows NRMSE ranges from 4.45% to 12.86% for the elevation values, and from 9.51% to 25.20% for the storage values. .

In this study, gauge observations reported by the Indian Central Electricity Authority (CEA, http://www.cea.nic.in/) were employed for validating. This dataset contains daily reservoir elevation, storage, and cumulative energy generation values for 30 hydropower reservoirs from 2008 to present (with a lag time of about 2-4 months).

Four reservoirs (Hirakud, Nagarjuna Sagar, Pong and Rengali), where observation data and training data were both available, were selected in this study to evaluate the performance of the developed algorithm. Further details about these reservoirs are provided in Appendix Table A1.

3.2.2. Methods

The new reservoir retrieval algorithm contains three steps, as summarized in the flowchart shown in Figure 3.1. First, a weighted horizontal ratio (*WHR*) based on the AMSR-E brightness temperatures is introduced for estimating the reservoir water surface area. Second, combined with the *A*-*H* relationship values from the training dataset (or from other sources), the *WHR* based surface area estimations are used to calculate the reservoir storage at a 4-day time step. Last, a Kalman Filter is applied to the 4-day results for reducing the high frequency noise and improving the accuracy of the reservoir storage estimations. A detailed explanation about each step is provided as follows.

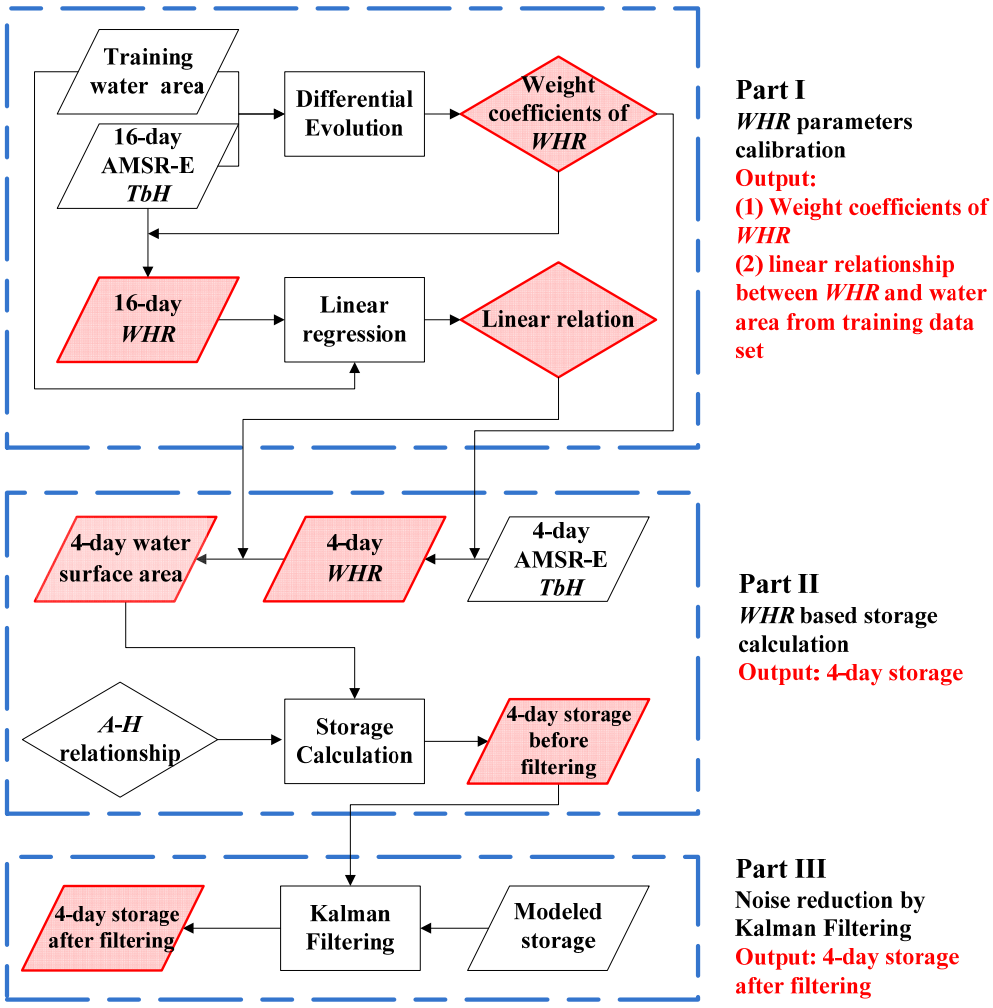


Figure 3.1 Flow chart of the algorithm. Trapezoids represent input/output data, diamonds represent formulas or coefficients, and rectangles stand for calculations. Data, formulas, and coefficients shaded in red are generated by the algorithm, while the unshaded elements are inputs from other sources.

3.2.2.1. Calibration of the weight coefficients of *WHR*

The brightness temperature of each pixel is not only influenced by the surface water area, but is also affected by physical temperature, soil wetness, and vegetation effects (e.g., emission, scattering). The *WHR*—which is first introduced in this study—can be solely dependent on the total surface water area, if desired, by adjusting the weight

coefficients. The weight coefficients in the *WHR* equation were calibrated by maximizing the coefficient of determination (R^2) between the *WHR* and the reservoir area from the training dataset. For each reservoir, AMSR-E pixels (both over the water body, and in the vicinity of the water body) were divided into target pixels (T_i) and surrounding pixels (S_j), where i and j are the indices of these two types of pixels. The target pixels are those pixels which can be at least partially covered by water when the reservoir is at capacity, while the surrounding pixels are those with no open water but are connected with one or more target pixels. When two pixels are "connected", it means they share at least one edge or one corner. Figure 3.2 uses the Hirakud reservoir as an example to illustrate the selection of the target pixels and surrounding pixels. The target and surrounding pixels for the other reservoirs in this study are shown in Appendix Figure A1.

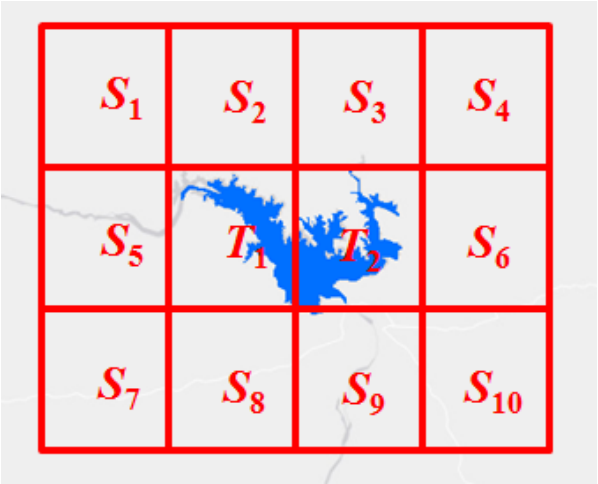


Figure 3.2 The target and surrounding pixels for the Hirakud reservoir.

Defined as the ratio between the weighted brightness temperatures from the target pixels and the surrounding pixels, the *WHR* is expressed below in Equation (3.1):

$$WHR = \frac{\sum_{i=1}^{N_T} W_i \times TbH(T_i)}{\sum_{j=1}^{N_S} W_j \times TbH(S_j)} \qquad (3.1)$$

where N_T and N_S denote the total number of target pixels (T_i) and surrounding pixels (S_j), respectively. *TbH* is the horizontally polarized brightness temperature, which is more sensitive to surface wetness than the vertically polarized brightness temperature. The calibration of the weight coefficients (W_i and W_j) against the training water area from MODIS (during the cloud free periods) aims to ensure that the *WHR* captures the best combination of *TbH* values from the target and surrounding pixels. The use of the *WHR* fully leverages the similarities and differences of the passive microwave emissions among the target pixel and surrounding pixels. Given the low spatial resolution of the AMSR-E observations, almost all of the target pixels are only partially covered by water. On the one hand, the emissivity of the open water portion is close to constant—and it is much lower than the non-water portion (i.e., bare soil or vegetated land [Gao et al., 2008]). On the other hand, the emissivity of the non-water portion depends on both the physical characteristics (e.g., soil texture and vegetation structure) and the surface wetness conditions (i.e., soil moisture). Together, the effective emissivity of a target pixel varies according to the water fraction and the soil moisture. It is reasonable to assume that the emissivity from some of the surrounding pixels is very close to the emissivity of the non-water portion from some of the target pixels. By using observations from the surrounding pixels as the denominator in Equation (3.1), there

exists a best combination of W_i and W_j such that the emission effect from the non-water portion within the target pixels (the numerator) can be illuminated from the *WHR*.

In this study, the Differential Evolution (DE) optimization algorithm was employed to calibrate W_i and W_j. The R^2 between the water surface area from the training dataset and the *WHR* was used as the objective function. The DE algorithm, developed by Storn and Price, [1997], is a floating-point encoded evolutionary algorithm for identifying the global maximum (or minimum) of an objective function. It searches for the global optimal solution by iteratively searching for, and improving upon, the candidate solutions. The DE algorithm has demonstrated good convergence properties, and has been applied in many different fields [Hejazi et al., 2011; Rekanos, 2008; Rocca et al., 2011; Zhong et al., 2013]. More detailed information about the DE algorithm is provided in the supplementary material. In the process of calibrating the weight coefficients, *TbH* was averaged every 16 days in order to match the temporal resolution of the water surface area from the training dataset (i.e., the MODIS-based area). Then a linear relationship between the *WHR* and the 16-day water surface area from the training dataset was generated for the reservoir of interest.

3.2.2.2. Reservoir surface area and storage calculation

Although calibrated using the 16-day data, the weight coefficients are independent of temporal resolution and are, in theory, applicable to the remotely sensed *TbH* at any time step. By applying the weight coefficients to the 4-day averaged *TbH* values, the *WHR* at a 4-day time step was obtained. Water surface area was then estimated based on

the *WHR* using the linear regression relationship from Section 3.2.2.1. Storage was therefore calculated according to Equation (3.2).

$$V_{RS}=V_c-(A_c+A_{RS})(h_c-h_{RS})/2 \quad (3.2)$$

where V_c, A_c, and h_c represent reservoir storage, surface area, and water elevation at capacity. V_{RS}, and A_{RS} are the 4-day *WHR* based storage and area, respectively. In addition, h_{RS} is the water surface elevation, which is regressed from the A_{RS} using the *A-H* relationship.

3.2.2.3. Kalman Filter

Because the reservoir storage data contained lots of noise at the 4-day time scale, an optimal sequential data fusion method—Kalman Filtering [Gelb, 1974]—was applied for noise reduction. The Kalman Filter, which works recursively, can run in near real time using only the present input measurements and the previously calculated state (and its uncertainty). It has been adopted by many applications in the field of hydrology, such as hydrological modeling [Crow and Van den Berg, 2010; Pan and Wood, 2006; Reichle et al., 2002], multi-source data fusion [Schwatke et al., 2015], and noise reduction [Batt and Carpenter, 2012].

In this study, Kalman Filtering was performed to reduce the noise contained in the 4-day AMSR-E based reservoir storage dataset by fusing a synthetic annual cycle reservoir storage model at the same temporal resolution. This synthetic model was first constructed based on the mean annual cycle of storage variations from the 16-day training dataset. It was then interpolated from 16-day to 4-day to match the remote sensing data. Finally, the optimal value of the reservoir storage (at the current time step)

was estimated using a Kalman Filter based on the optimal value of the previous time step, the 4-day AMSR-E based storage value, and the modeled storage at the current step. More detailed information about the filtering algorithm is provided in the supplementary material.

3.3. Results

Since the storage estimating skill depends largely on whether the *WHR* can capture the variations of water surface area effectively, calibration of the weight coefficients in the *WHR* is a key component. The effectiveness of the *WHR* approach was examined in Section 3.1 at a 16-day temporal resolution by comparing the *WHR* based water surface areas with the training dataset. In Section 3.3.2, the time series of storage variations (both before and after filtering) were compared with *in situ* gauge observations to test the skill of Kalman Filtering for reducing noise.

3.3.1. Effectiveness of the WHR based area estimation approach

The 16-day *WHR* based water surface areas were compared with the training dataset values over the four studied reservoirs from 2003 to 2007 in Figure 3.3. The R^2 between the training dataset and the *WHR* based area ranges from 0.62 to 0.85. This suggests that the weight coefficients of *WHR* are robust and that the *WHR* is a good indicator of reservoir area. Among all of the tested reservoirs, the Hirakud reservoir has the highest R^2. The land cover types in the target and surrounding pixels for Hirakud are similar with each other, which makes the *WHR* more representative of open water. The good performance of the weight coefficients in the Hirakud reservoir is also attributed to its high quality training data (as shown in Table S1). However, it should be stated that an

even higher accuracy of the training data would further improve the performance of the calibration, and lead to a better R^2 value.

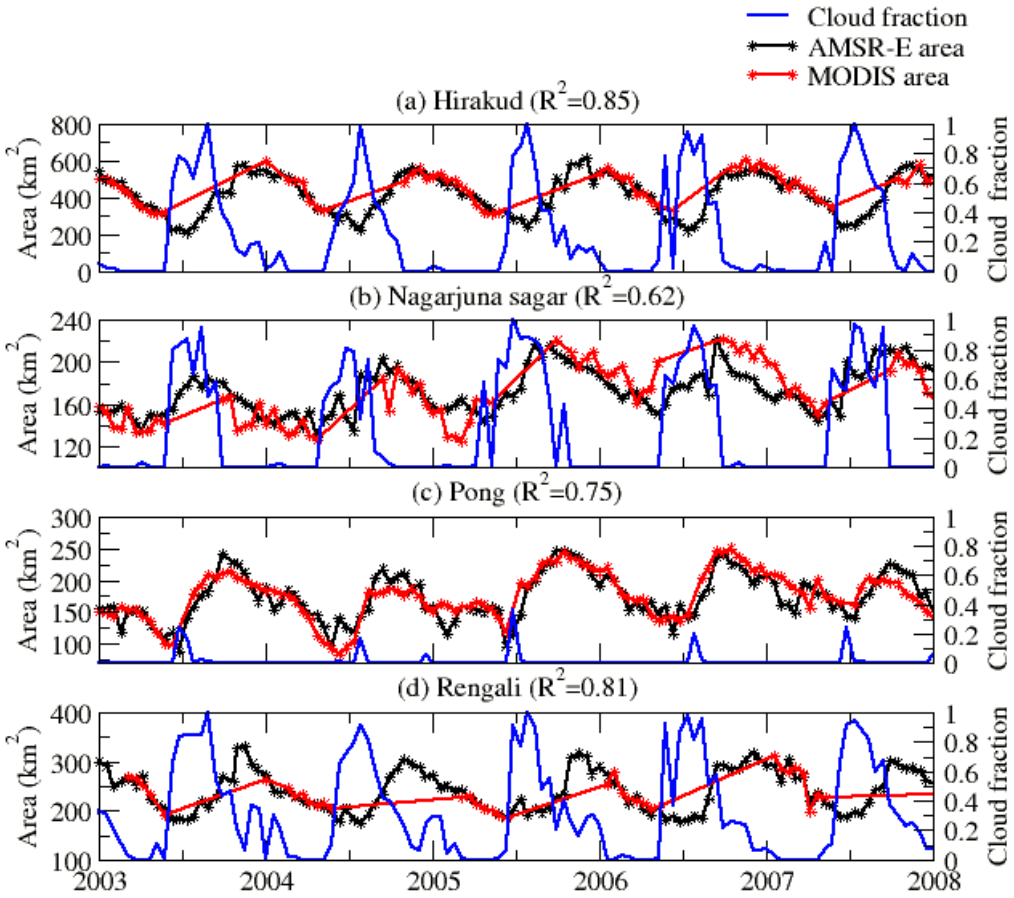


Figure 3.3 Comparisons of water surface area estimated by the training dataset and the *WHR* during the calibration period (from 2003 to 2007) for (a) Hirakud; (b) Nagarjuna Sagar; (c) Pong; and (d) Renglai reservoir. Cloud coverage from MOD13Q1 was utilized to show the limitations with MODIS measurements.

In addition, because the Hirakud reservoir has the largest surface area among all the studied reservoirs, it is easier to capture the area variations (by using microwave data with low spatial resolution) than it is for reservoirs with small surface areas. The inferior

performance of the Nagarjuna Sagar reservoir (R^2=0.62) is attributed to the relatively low accuracy of its training data.

The AMSR-E based area has demonstrated a clear advantage over the MODIS based training area by providing results under all-weather conditions, even with severe cloud coverage (Figure 3.3). During the 5-year period from 2003 to 2007, the AMSR-E instrument collected 95%, 55%, 6%, and 219% more points at a 16-day step than MODIS over the Hirakud, Nagarjuna Sagar, Pong and Rengali reservoirs, respectively. This additional information is especially meaningful for flood monitoring and forecasting.

3.3.2. Validation of the AMSR-E based 4-day storage estimations

The AMSR-E based reservoir storage estimations, both before and after noise reduction, were validated comprehensively by comparing them with the *in situ* observations from 2008 to 2011. Figure 3.4a shows the validation results before noise reduction. Among the four reservoirs, the Hirakud and Pong reservoirs have the highest R^2 values (0.61 and 0.67). Again, the lowest R^2 was found at Nagarjuna Sagar, which is attributed to a weak performance of *WHR* (as discussed in Section 3.3.1). The NRMSE values are 30%, 47%, 30% and 49% for the Hirakud, Nagarjuna Sagar, Pong, and Rengali reservoirs, respectively.

The 4-day storage time series, after noise reduction through Kalman Filtering, were compared with *in situ* data in Figure 4b. The comparison suggests that the filtered storage time series are much improved. The R^2 values are 0.63, 0.41, 0.74, and 0.63 for the Hirakud, Nagarjuna Sagar, Pong, and Rengali reservoirs. The NRSMEs are 28%,

42%, 29% and 34%, respectively. Although AMSR-E has the ability to penetrate clouds, the AMSR–E time series data is still affected by noise introduced by severe cloud coverage or rain events [Draper et al., 2009]. Generally, more noise contained in pre-filtered data provides greater room to improve (through the Kalman Filtering). For example, the most significant improvement in the present study is achieved at the Rengali reservoir, which has the largest fluctuations and the longest cloud coverage period. This suggests that the pre-filtered Rengali is the most affected by noise. In contrast, the R^2 values changed little for the Hirakud and Pong reservoirs, where the pre-filtered data were already relatively smooth.

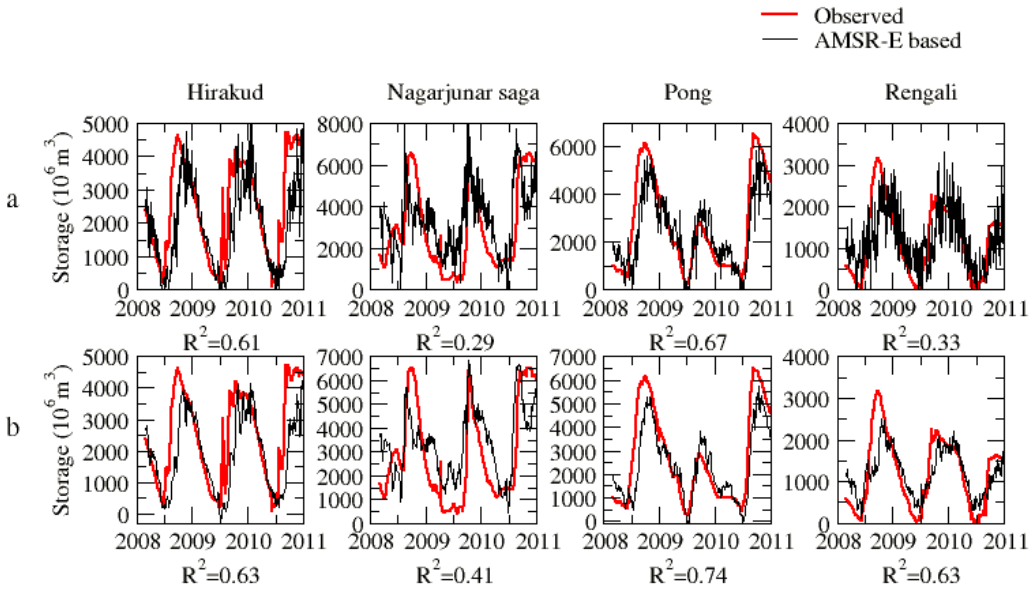


Figure 3.4 Validation results for AMSR-E based water storage at a 4-day time step. a) before Kalman Filtering; and b) after Kalman Filtering.

3.4. Conclusions and discussion

In this study, a new method was developed to estimate 4-day reservoir storage variations under all-weather conditions by leveraging cloud-penetrating passive microwave observations. A previous dataset, solely derived from MODIS and ICESat measurements, was utilized as a training dataset in this method for three purposes. First, the training dataset was used to calibrate the weight coefficients from a passive microwave brightness temperature index (i.e., *WHR*) at 36.5 GHz. Second, the *A*-*H* relationship derived from the training dataset was employed for estimating both the reservoir surface area and the storage from the (AMSR-E observation based) *WHR*. Third, the noise associated with the 4-day reservoir estimations was further reduced by fusing with the storage from a training dataset based annual cycle model through a Kalman Filter. Validation results against *in situ* observations over the four testing reservoirs in South Asia suggest that the filtered AMSR-E 4-day storage estimations have a relatively high accuracy level (with R^2 values ranging from 0.41 to 0.74).

This method makes new contributions to reservoir remote sensing in several ways. First, it is the first time that passive microwave remote sensing data have been fused with other satellite data for quantifying the area and storage variations of individual reservoirs. The combined use of multi-pixel data and the DE optimization algorithm (for calibrating the *WHR* weight coefficients) has alleviated the limitation resulting from AMSR-E's low spatial resolution. Second, the capability of measuring reservoirs at a 4-day time step under all-weather conditions is by far the highest temporal resolution that has been achieved in satellite remote sensing of inland water bodies (without combining

observations from multiple radar altimeters). In this, the centimeter wavelength radiation of passive microwave is essential for penetrating clouds, while the Kalman Filter has enhanced the data quality. This capability is expected to improve flood monitoring and water management significantly. Third, it is worth noting that not only water surface area—but also the water surface elevation, and the water storage values—can be used for calibrating the coefficients of the *WHR*. This suggests that the algorithm has great flexibility to work with many other past, current, and future sensors—and/or data directly from *in situ* measurements.

This algorithm still has some limitations which need to be discussed. Firstly, the accuracy of the proposed algorithm depends on the quality of the training dataset. For certain reservoirs which have complex shapes (such as too narrow or meandering), the area/storage variations are difficult for MODIS to capture [Gao et al., 2012]. In these cases, the performance of the algorithm (which uses MODIS water surface area as training data) will be limited. Secondly, because the performance of this algorithm depends on the R^2 between the AMSR-E based *WHR* and the MODIS based training data, it is suggested that this algorithm should not be used if the R^2 is less than 0.5. According to our analysis of the 21 reservoirs in the training dataset, the standard deviation of the MODIS based area needs to be larger than 20 km^2 to meet the above criteria for R^2.

This algorithm can result in reservoir storage products which support water management in long term (climatological) observations and short term monitoring. For instance, given the long term availability of passive microwave remote sensing data—such as that from the Special Sensor Microwave Imager (SSM/I) and the Special Sensor

Microwave Imager Sounder (SSMIS) on board the Defense Meteorological Satellite Program (DMSP) satellites—a climatological reservoir record can be generated. By leveraging new data from the Soil Moisture Active Passive (SMAP) satellite mission launched in 2015, near real time reservoir storage monitoring can be achieved. This algorithm may also contribute to future satellite missions such as the SWOT, which will provide a direct water surface measurement for about two-thirds of global lake and reservoir storage (with an area > 0.06 km^2). Utilizing the storage estimations from ICESat and MODIS as the training dataset, many more global reservoirs can be monitored using passive microwave remote sensing at a high temporal resolution.

CHAPTER IV

MONITORING RESERVOIR STORAGE VARIATIONS IN SOUTH ASIA FROM SATELLITE IMAGERIES AND DIGITAL ELEVATION MODEL

4.1. Introduction

Reservoirs, which are managed by storing and releasing water under pre-determined operation rules, play an important role in mitigating floods and improving the efficiency of the water supply for municipal, industrial, and agricultural demands [Bai et al., 2015; Fallah-Mehdipour et al., 2012; Haddeland et al., 2014; Zhao et al., 2016]. Although most (if not all) human operated reservoirs are monitored in realtime, such reservoir storage information is not commonly shared among countries located within international river basins, especially among countries with conflicting interests. The difficulties with regards to information sharing limit the effectiveness of reservoir management. For instance, false alarms are likely to occur in the current global flood monitoring systems, in the flood control functions of reservoirs are not taken into consideration [Wu et al., 2012]. In addition, the lack of reservoir storage are likely to reduce the reliability of drought analyzing systems [Pulwarty and Sivakumar, 2014; Vicente-Serrano et al., 2012].

Free of the data sharing limitations, remote sensing technology provides a promising alternatives for monitoring the reservoirs from space [Gao et al., 2012; Lettenmaier et al., 2015; Rodrigues et al., 2012]. Reservoir storage information is

usually inferred from remotely sensed water surface area estimations and/or water surface elevation [Gao, 2015].

For measuring water surface area, the common approach is to estimate the water extent thorough classifying the optical satellite imageries such as the those from Landsat or MODIS. With a spatial resolution of 30 m, Landsat observations are made over the same region every 16 days. Despite Landsat's high spatial resolution, its low temporal resolution can be exacerbated by cloud contamination. This discontinuity of Landsat greatly hampers its capability for reservoir monitoring. Compared with Landsat, the spatial resolution of the MODIS sensors (250 m/500 m/1000 m) is much coarser. But its high temporal resolution (i.e., daily) allows its 16-day composite to be less affected by cloud cover than that of the (once every 16 days) Landsat images. This makes MODIS a better tool than Landsat for monitoring reservoir area variations.

To calculate reservoir storage from remote sensing, the other critical variable is water surface elevation. Satellite radar altimeters, which measure the surface elevation by timing the pulses emitted from the sensor and the echo reflected by the surface, have been utilized in many studies [Calmant et al., 2008]. The most commonly used radar altimeters, such as ENVISAT and T/P satellites, can only be used for large lakes and reservoirs due to the footprint size (~10 kilometers) [Gao, 2015]. In addition to the radar altimeters, the GLAS onboard the ICESat has a higher horizontal spatial resolution (approximately 70 m), has been utilized to monitor relatively small lakes/ reservoirs [Wang et al., 2013; Zhang et al., 2011; Zhang and Gao, 2016]. use of ICESat/GLAS for monitoring water elevations operationally has been hindered by its very low temporal

resolution (91 days). As a result, ICESat applications usually focus on interannual elevation variations of natural lakes [Phan et al., 2012].

There have been some studies focusing on monitoring reservoir storage by leveraging information from multiple satellites, but the density of the spatial coverage is still limited. For instance, Liebe et al., [2005] calculated lake volumes as a function of water surface area, with the relationship between these two based on surveyed bathymetry survey. Smith and Pavelsky, [2009] created a *V-A* relationship through gauge measured lake storage and remotely sensed water surface area so that lake storage variations can be estimated by monitoring water surface area from space. These approaches are relatively straightforward, but the requirements for bathymetry or gauge measurements still inhibit these approaches' applicability for regions where such information is inaccessible. For monitoring reservoirs solely through satellite remote sensing, both water surface elevation and water surface area should be collected from space. Through a study that was based on a combination of water surface area from MODIS and reservoir elevations from radar altimetry [Gao et al., 2012], a global reservoir storage dataset was generated by using remote sensing data solely. However, only about 15% of global reservoirs (by storage capacity) can be monitored from space primarily because radar altimeters are of very low spatial resolution (several kilometers). By using ICESat to replace radar altimeters, a dataset was established representing 28% of the total capacity in South Asia [Zhang et al., 2014]. In spite of such progress, the reservoir observation network is still too sparse to support operational applications due

to the limited spatial coverage of ICESat. Therefore, the lack of dense spatial representation of altimeters still remains an issue.

South Asia, which contains one of the largest and densest populations in the world, is very prone to flooding. According to the statistics records, South Asia experiences one of the highest fatality rates caused by floods [Adhikari et al., 2010]. The failure with regard to communicating reservoir storage and management information further exacerbates the casualties and economic losses from flood events. Meanwhile, the currently available remotely sensed reservoir storage datasets cannot offer high-density spatial coverage. For instance, radar altimetry data is only available for six reservoirs in this region which accounts for 10.70% of the total capacity in South Asia (according to GOSH data http://www.legos.obs-mip.fr/soa/hydrologie/hydroweb/Page_2.html and USDA reservoir data set https://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/). The ICESat elevation data covers only around 28% of South Asian reservoirs [Zhang et al., 2014]. Given the strong societal needs, it is critical to acquire reservoir storage information with a large spatial coverage for minimizing vulnerabilities and maximizing benefits through good reservoir management.

In order to extend the spatial representativeness of the remote sensing reservoir storage dataset, a new algorithm is developed by leveraging the global coverage capability of the DEM collected by the SRTM in February, 2000. SRTM provided the DEM over land at a 30 meter resolution between 60 N and 54 S, with a relative vertical accuracy of ~6m and an absolute accuracy of ~16 m [Rabus et al., 2003]. It has been

used widely in the monitoring of land surface water resources, such as studies estimating the glacier variations [Berthier et al., 2006; Surazakov and Aizen, 2006] and surface water storage change [Papa et al., 2013]. Because of its high consistency (with regards to accuracy) and its global coverage [Rabus et al., 2003; Surazakov and Aizen, 2006], the SRTM DEM was used to extract the *A*-*H* relationship for reservoir storage calculations in this study.

The overarching goal of this chapter is to improve upon the spatial coverage of our current remotely sensed reservoir storage dataset in the South Asia region. The MODIS based water s area estimation algorithm from [Zhang et al. 2014] was utilized such that high quality water storage estimations could be achieved using water surface area estimations and the water surface area – height (*A*-*H*) relationship derived from SRTM DEM. In addition to the data analysis and results validations, storage estimation uncertainties (due to reservoir surface area retrieval algorithm parameterization and elevation measurement errors) were also quantified.

Table 4.1 Detail Information for the 27 Reservoirs

I. D.	Reservoir	Country	Location (° N , ° E)	Area (km^2)	Capacity (10^6 m^3)	Purpose[a]	*A*-*H* relationship[b]
01	Almatti	India	16.33, 75.89	424	2631	E	y=0.026x+507.17
02	Bango	India	22.61,82.60	104	3416	I,E	y=0.201x+332.57
03	Bargi	India	22.95,79.93	268	3920	I,E	y=0.104x+400.28
04	Chandil	India	22.98,86.02	139	1961	I,E	y=0.166x+170.15
05	Gandhi Sagar	India	24.71, 75.55	578	5600	E	y= 0.034x+378.24

Table 4.1 (continued)

I. D.	Reservoir	Country	Location (° N , ° E)	Area (km^2)	Capacity (10^6 m^3)	Purpose[a]	*A-H* relationship[b]
06	Hirakud	India	21.52,83.85	603	4079	I,E	y=0.270x+174.48
07	Karnafuli	Bangladesh	22.5,92.23	777	6477	I,E,F	y=0.024 x +23.375
08	Krisharaja Sagar	India	12.42,76.57	100	1369	I,E,W	y=0.134 x +736.91
09	Linganamakki	India	14.18,74.85	316	4178	E	y=0.079x+542.95
10	Mangla	Pakistan	33.13,73.64	251	7300	I,E,F	y=0.166 x +319.61
11	Malaprabha	India	15.82,75.09	130	1068	I,E	y=0.136 x +619.53
12	Matatila	India	25.10,78.37	139	1133	I,E	y=0.095 x +292.84
13	Nagarjuna Sagar	India	16.57,79.31	240	6538	I,E	y=0.270 x +118.8
14	Narayanapura	India	16.22,76.35	102	1071	I	y=0.105x+482.91
15	Pong	India	31.97,75.95	260	6946	I,E	y=0.212 x +366.98
16	Rajghat	India	24.76,78.23	224	2172	I,E	y=0.070 x +350.35
17	Ranjit Sagar	India	32.44,75.73	56	2200	E	y=1.284x+441.10
18	Rengali	India	21.28,85.03	392	3168	I	y=0.070 x +100.88
19	Rihand	India	24.20,83.01	485	5846	I,E	y=0.083 x +232.99
20	R. P. Sagar	India	24.92,75.58	210	1568	I,E	y=0.123 x +325.49
21	Singur	India	17.75,77.93	129	850	W	y=0.053 x +517.21
22	Srisailam	India	16.09,78.90	560	7105	I,E	y=0.042 x +254.05
23	Supa	India	15.28,74.53	120	4178	E	y=0.460 x +506.89
24	Tawa	India	22.56,77.98	200	2310	I	y=0.117 x +338.36
25	Tungabhadra	India	15.27,76.33	390	3764	I,E	y=0.052 x +483.92
26	Ukai	India	21.25,73.59	512	6199	I,E,F	y=0.042 x +81.364
27	Yeldari	India	19.72,76.73	82	934.3	I,E	y=0.223 x +443.45

[a]I is irrigation, E is electricity generation, W is water supply, and F is flood control.
[b]y is water surface height and x is area.

4.2. Data

4.2.1. Remote sensing data

In this study, the two main remote sensing datasets are the STRM DEM and the MODIS imageries. The DEM was used for inferring the area-elevation (*A*-*H*) relationship. The MODIS imageries were used to derive surface area estimations, which were applied to the *A*-*H* relationship to generate a long term time series of reservoir storage. The DEM from SRTM provides land surface elevation data at a 30-m spatial resolution. The DEM data were collected by SRTM during an 11-day mission in February 2000, covering a near-global domain from 56° S to 60° N [Farr et al., 2007]. The reservoir surface area was calculated from the MODIS/Terra 16-day 250-m resolution vegetation indices product (MOD13Q1). Specifically, an image classification algorithm (see Section 3.2.1) was applied to the Normalized Difference Vegetation Index (NDVI) imageries to extract the reservoir area. From 2000 to 2015, a total of 365 imageries were processed. Both datasets were obtained from the Land Processes Distributed Active Archive Center (LPDAAC) website.

In addition, the water surface elevation data from ICESat was used to evaluate the accuracy of the *A*-*H* relationship derived by SRTM. The ICESat mission from 2003 to 2010 provided measurement of ice, cloud, and aerosol heights and topography. Obtained through the U.S. National Snow and Ice Data Center (NSIDC, http://nsidc.org/data/icesat/), the ICESat data was with high vertical precision (~10cm) [Phan et al., 2012; Wang et al., 2013].

4.2.2. Gauge observation data

Gauge observations released by the Indian Central Electricity Authority (CEA, http://www.cea.nic.in) were used to validate the remotely sensed reservoir storage dataset. This gauge data contain daily reservoir water level and storage data for 30 hydropower reservoirs from 2013 to present, at near real time. In addition, we also have the record from 2008 to 2011 (in the same format) downloaded from this same source in May, 2014.

4.3. Reservoir selection and methodology

4.3.1. Reservoir selection

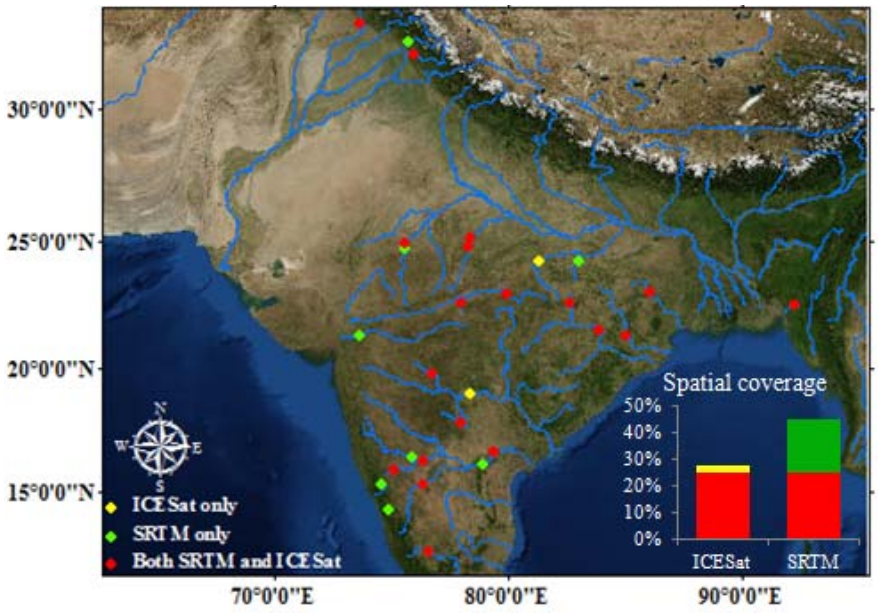


Figure 4.1 Locations of 27 reservoirs which can be monitored by remote sensing. Yellow dots represent reservoirs that can only be monitored using the MODIS-ICESat approach. Green dots are reservoirs which can only be monitored through MODIS-STRM. Red points are reservoirs which can be monitored using both approaches. For each reservoir, detail information is provided in Table 4.1.

Reservoirs in this study were selected from the GRanD database. Two criteria were used to identify the selected reservoirs. First, the reservoir maximum area at capacity had to be larger than 55 km^2. This was used to guarantee that the surface area can be estimated with high accuracy using medium resolution MOD13Q1 NDVI imageries. Second, the reservoir surface area from the SRTM DEM had to be less than 60% of its maximum surface area—otherwise the ranges of area and elevation detected by SRTM DEM would be too small to infer the *A-H* relationship accurately. Following the above criteria, a total of 27 reservoirs were chosen for this study, which accounts for 45% of the South Asian reservoir capacity. Figure 4.1 shows the locations of these reservoirs, and it also compares the reservoirs from this study with those from Zhang et al. [2014] and from radar altimetry sensing.

4.3.2. Methodology for reservoir storage estimation

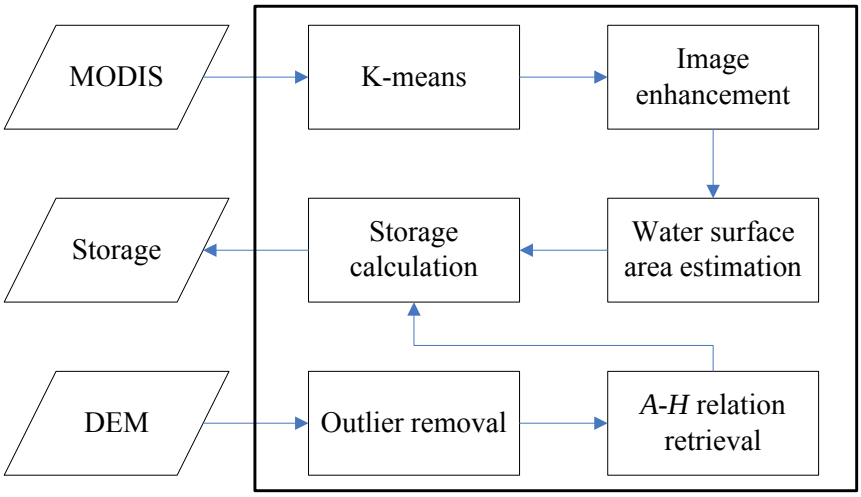


Figure 4.2 Flow chart of the MODIS-SRTM based reservoir storage estimation algorithm.

The MODIS-SRTM based reservoir storage estimation algorithm is illustrated using the flowchart in Figure 4.2. The overall workflow of this algorithm (referred to as the "MODIS-SRTM approach" thereafter) contains three major steps. First, water surface area is estimated from MODIS NDVI imageries via an enhanced classification procedure. Second, the *A–H* relationship is generated from the DEM information by regressing the cumulative area values against their corresponding elevation values within the delineated reservoir maximum domain. Third, by applying the water surface area estimations to the *A–H* relationship, the reservoir storage variations are calculated. Further details of these steps are provided as follows.

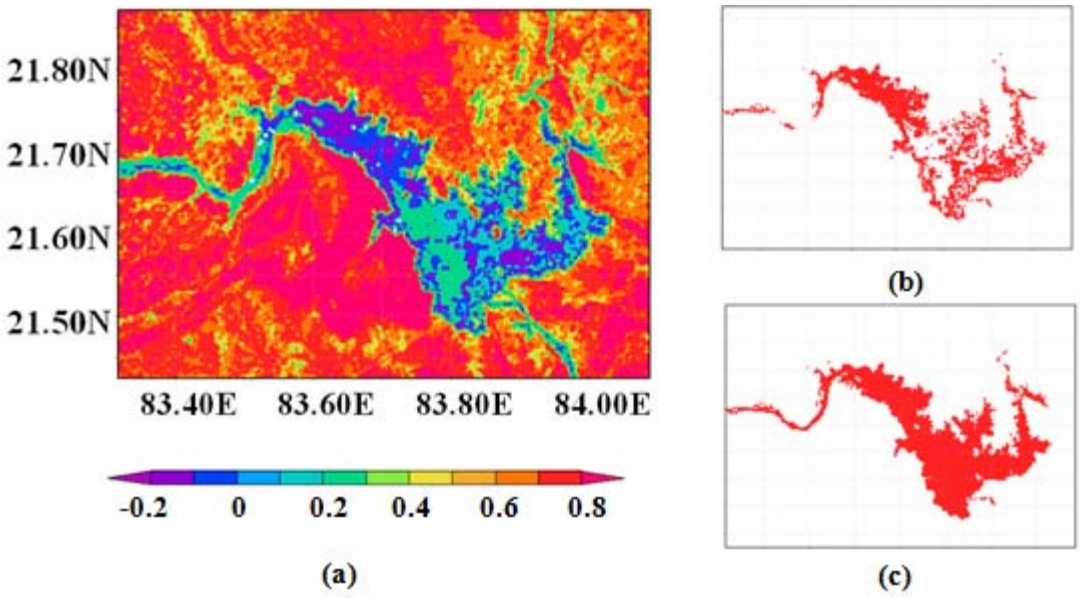


Figure 4.3 Hirakud reservoir area using the MODIS NDVI based classification reprinted from [Gao et al., 2016] (a) the original MODIS NDVI image of day 273 of 2005; (b) the classification result without image enhancement; and (c) the classification result after image enhancement.

4.3.2.1. Surface area estimation

The reservoir water surface area was estimated using the enhanced K-means classification approach by Zhang et al. [2014].

First, a threshold of 0.1 was applied to each 16-day MODIS NDVI image from 2000 to 2015 (where pixels with NDVI values less than 0.1 are considered as water). Based on these simplified classification results, a mask image was created to represent the water coverage percentile and to delineate the domain of reservoir. Then, the K-means clustering algorithm [Jain, 2010] was utilized to identify all water pixels within the masked area of the MODIS NDVI images. Finally, a classification enhancement procedure was used to fine-tune the results from the previous step. The main idea of this last step is to use the percentile information from the mask image as a reference to assign an appropriate class to the misclassified pixels. Figure 4.3 shows a real example of how the water surface estimation can be improved when the K-means classification of the original NDVI image is of low quality.

4.3.2.2. Area-elevation (*A*-*H*) relationship development

The SRTM DEM data were used to extract the *A*-*H* relationship for each reservoir. As an approximation, the relationships for all reservoirs are assumed linear [Gao et al., 2016]. To capture the relationship, we first delineated the reservoir region from the DEM. A simplified example of a delineated reservoir from DEM is shown in Figure 4.4a. To assure that the *A*-*H* relationship realistically represents the variations of *H* when water surface area changes, the following assumption was made as a constraint: any pixel which was not directly connected to the main body of the reservoir of interest was

excluded as noise. After delineating the reservoir maximum coverage from the DEM, the cumulative area (e.g., A3) at a given elevation (e.g., H3) can be estimated by counting the number of pixels with elevations equal to or smaller than that elevation value (H3).

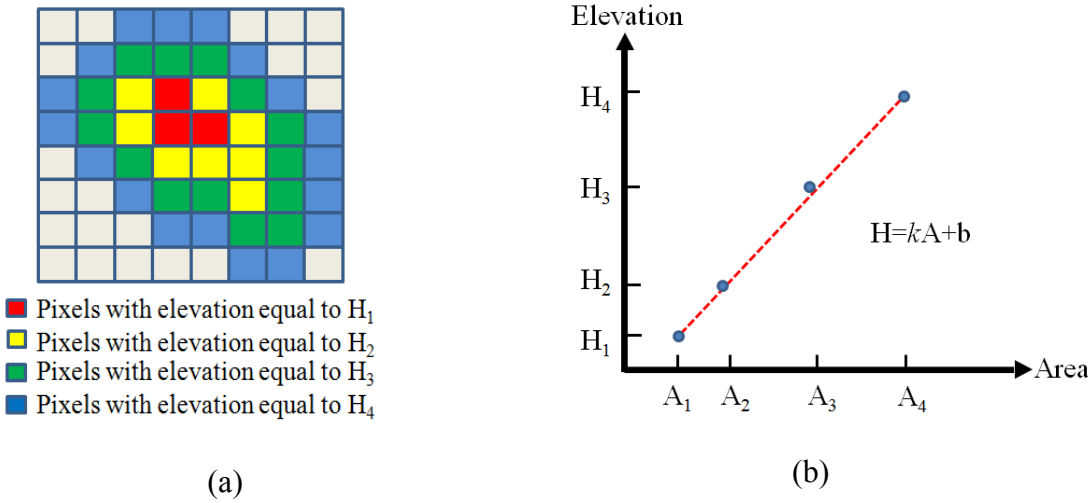
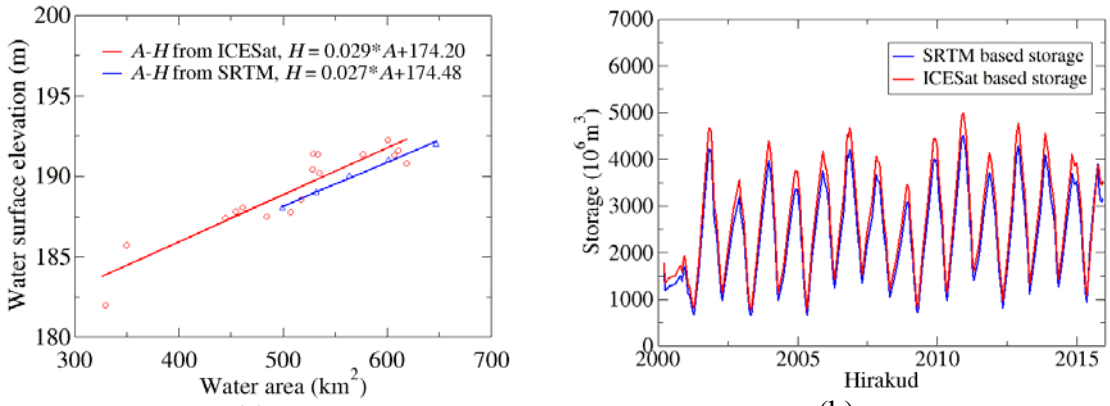


Figure 4.4 (a) A simplified example of a delineated reservoir from the SRTM DEM, where H1> H2> H3> H4; (b) the corresponding *A*-*H* relationship inferred from the simplified example in (a).

By regressing the cumulative area values against the elevation values, the *A*-*H* relationship for the reservoir of interest can be established (Figure 4.4b).

A real example of the development of the *A*-*H* relationship over the Hirakud reservoir is shown in Figure 4.5a.



(a) (b)

Figure 4.5 (a) The *A*-*H* relation developed from SRTM compared with the relation derived from ICESat; (b) time series of the storage estimation for the Hirakud reservoir from the SRTM based approach and the ICESat based approach.

In addition, this *A*-*H* relationship is compared with that derived from MODIS area and ICESat elevation for cross validation purposes. The MODIS-ICESat based *A*-*H* relationship was adopted from Zhang et al. [2014]. The *A*-*H* relationship from the ICESat based approach was capable of capturing more water surface elevation values because it had a longer working period. The SRTM based *A*-*H* relationship was derived from topographic points with a shorter range of elevation values. Because the *A*-*H* relationship is relatively constant when the elevation changes, the *A*-*H* relationships derived from ICESat and SRTM are similar with each other.

4.3.2.3. Storage estimation

According to the previous studies [Gao et al., 2012; Zhang and Gao, 2016; Zhang et al., 2014], reservoir storage can be estimated based on the remotely sensed water surface area and elevation values using equation (4.1):

$$V_{RS} = V_C - \left(A_C + A_{RS}\right)\left(H_C - H_{RS}\right)/2 \qquad (4.1)$$

where V_c, A_c, and H_C represent storage, area, and water elevation at capacity, and V_{RS}, A_{RS}, and H_{RS} are the remotely sensed storage, area, and water height at the monitoring time.

In this MODIS-SRTM approach, since H_{RS} can be calculated by applying the *A*-*H* relationship ($H=kA+b$) to the MODIS area estimation (i.e., A_{RS}), the reservoir storage value is calculated through equation (4.2), which is transformed from equation (4.1).

$$V_{RS} = V_C - \left(A_C + A_{RS}\right)\left(A_C - A_{RS}\right)k/2 \quad (4.2)$$

where k is the slope from the *A*-*H* relationship.

Using the methods explained in this section, the reservoir storage was calculated for the 27 selected South Asian reservoirs from 2000 to 2015. Using the Hirakud reservoir as an example, Figure 4.5b compares the time-series of reservoir storage from this MODIS-SRTM approach with that from the MODIS-ICESat approach by Zhang et al. [2014]. Results suggest that these two sets of storage estimations are in good agreement. However, compared with the MODIS-ICESat based approach, the storage values from this study tend to be underestimated owing to the different *A*-*H* relationships. In order to better understand the error statistics of the two approaches, validations using gauge data were carried out in Section 4.4.1.

4.4. Results

Figure 4.6 shows the time series (2000 to 2015) of the 27 selected reservoirs in South Asia, with an integrated capacity of 118.76 km^3 (as compared to the region's total capacity of 263.91 km^3). Compared with utilizing elevation information from ICESat,

the MODIS-SRTM approach has allowed for eight additional reservoirs to represent a 5.07 km^3 increase of the overall storage capacity (an equivalent of 17%) of South Asia.

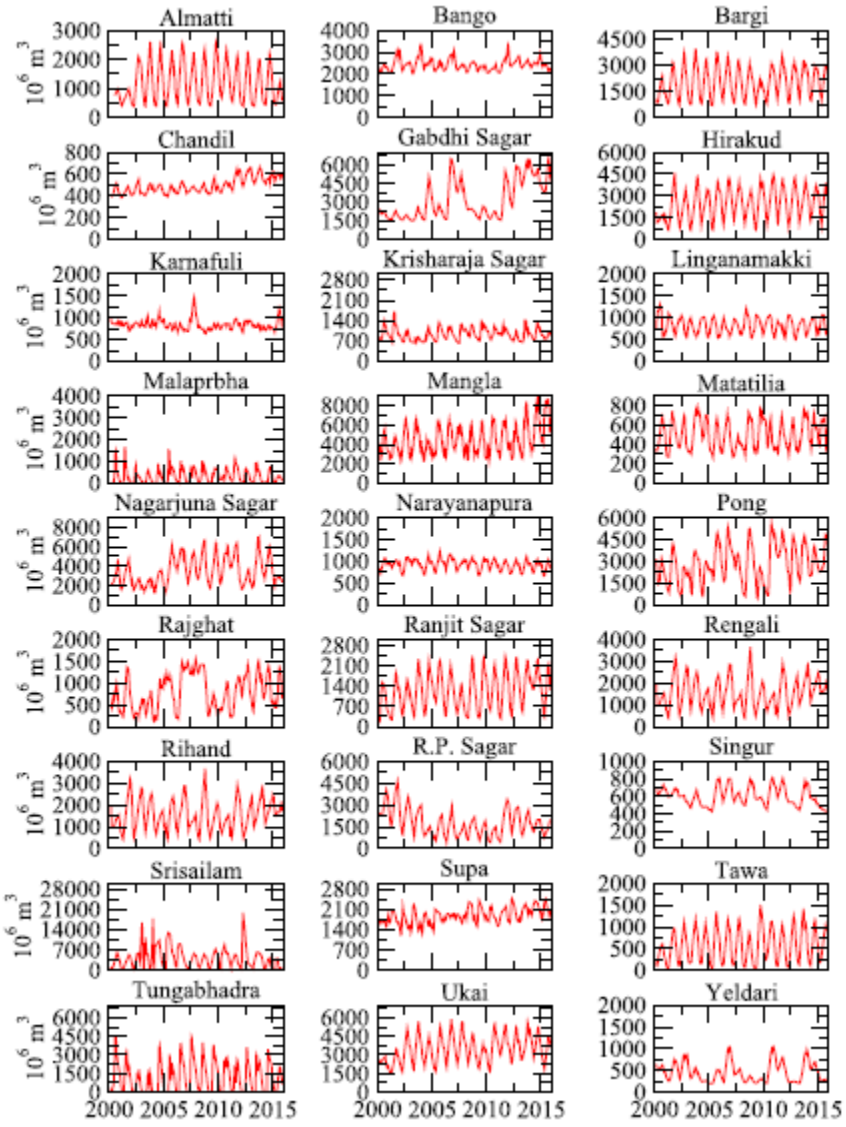


Figure 4.6 Remotely sensed storage time series of the South Asian reservoirs monitored in this study

4.4.1. Validation results

The MODIS-SRTM based reservoir storage was validated over 11 reservoirs where gauge observation data are available. Table 4.2 shows the validation results using three statistical criteria. These are the coefficient of determination R^2, the relative bias *B*, and the normalized root-mean square error NRMSE.

Table 4.2 Statistical Validation Results for the Remotely Sensed Reservoir Storage from the MODIS-SRTM Approach

I. D.	Reservoir name	R^2	Bias(%)	NRSME(%)
01	Almatti	0.84	12.40	35.87
05	Gabdhi Sagar	0.69	6.25	15.46
06	Hirakud	0.88	-12.25	18.14
14	Nagarjunasagar	0.82	2.41	27.58
15	Pong	0.88	20.32	27.76
17	Ranjit Sagar	0.47	17.77	37.69
18	Rengali	0.79	-13.51	15.62
19	Rihand	0.84	-16.22	28.69
20	R. P. Sagar	0.91	-3.24	24.61
22	Srisailam	0.90	-31.7	32.75
26	Ukai	0.81	-14.76	15.93

In Table 4.2, most of these results are highly correlated with the CEA gauge observations. The coefficients of determination (R^2) range from 0.47 to 0.91, with a mean value of 0.8. The lowest R^2 was found over the Ranjit Sagar reservoir. This reservoir has a relatively small area (56 km^2 at capacity) and it is very meandering (with a high shoreline to area ratio), making it difficult to accurately estimate the surface area

from the medium spatial resolution MODIS data [Gao et al., 2012]. Using multi- criteria evaluation, a more comprehensive evaluation is given. The correlation is more affected by the surface area estimation from MODIS, while the NRMSE and *B* are more affected by the combined effect from the water surface area estimation and the *A*-*H* relationship. For instance, using the Srisailam reservoir as an example, the correlation of determination is the second highest among all of the validated reservoirs—but the NRSME is relatively large due to the inaccurate *A*-*H* relationship inferred from SRTM DEM for this reservoir. Another example is the Ranjit Sagar reservoir. Here even the surface area estimation has a large error, and the accuracy of calculated storage has been improved by a more accurate *A*-*H* relationship.

4.4.2. Comparison with the ICESat based approach

In order to better understand the characteristics of the new proposed approach, the MODIS-SRTM results were compared comprehensively with results from the MODIS-ICESat based approach (in terms of their spatial coverage and validation results against gauge observations).

4.4.2.1. Spatial coverage

With the full coverage 2-diminsional elevation data at a fine spatial resolution of 30 m, the DEM approach covers eight additional reservoirs which cannot be monitored by ICESat (shown by Figure 4.1). There are only two reservoirs for which ICESat data are accessible but the DEM approach failed at developing the *A*-*H* relationship needed for estimating the reservoir storage. One is the Bansagar reservoir constructed in 2006, which was non-existing when the SRTM DEM measurements were collected (2000).

The other reservoir is Sriram Sagar, which was almost at its full level during the SRTM flight time. Nonetheless, the total storage monitored using the DEM based approach represents 45 % of the total capacity in South Asia, which is 17% more than the ICESat based approach. As a result, storage variations over multiple reservoirs within a river basin can be assessed, which is essential for regional water management purposes. For instance, with these new reservoirs included in the data set, the total storage of reservoirs in Krishna river basin has increased from 2983 to 3895 km^3. The Krishna River is the fourth-largest river in terms of water inflow and river basin area in India. Therefore, the increased storage values from these new reservoirs has significantly strengthened the spatial representativeness of reservoirs in this river basin. Another example is the Ukai Dam across the Tapti river, which was constructed for the purposes of irrigation, power generation and flood control. The Tapti river basin accounts for nearly two percent of the total area of India. However, before leveraging SRTM to monitor the reservoir storage, the previous remote sensing reservoir dataset, which was derived from MODIS-ICESat approach, contains no reservoir in this basin. In August 2006, a flooding event happened in the surrounding region which caused eight villages to be inundated with water. To mitigate the damage from the flood, the outflow of the reservoir was reduced which caused it to reach its maximum water level–which agrees with the remote sensing results in Figure 4.6 (http://www.oneindia.com/2006/08/09/floods-paralyse-surat-city-outflow-reduced-from-ukai-dam-1155117864.html).

4.4.2.2. Comparison of storage validation results from the SRTM based and ICESat based approaches

The remotely sensed reservoir storage data, both from MODIS-SRTM and MODIS –ICESat approaches, were validated over five reservoirs (Hirakud, Nagarjuna Sagar, Pong, and Rengali, R. P. Sagar) where gauge observations and *A*-*H* values (from both MODIS-ICESat and SRTM) were available.

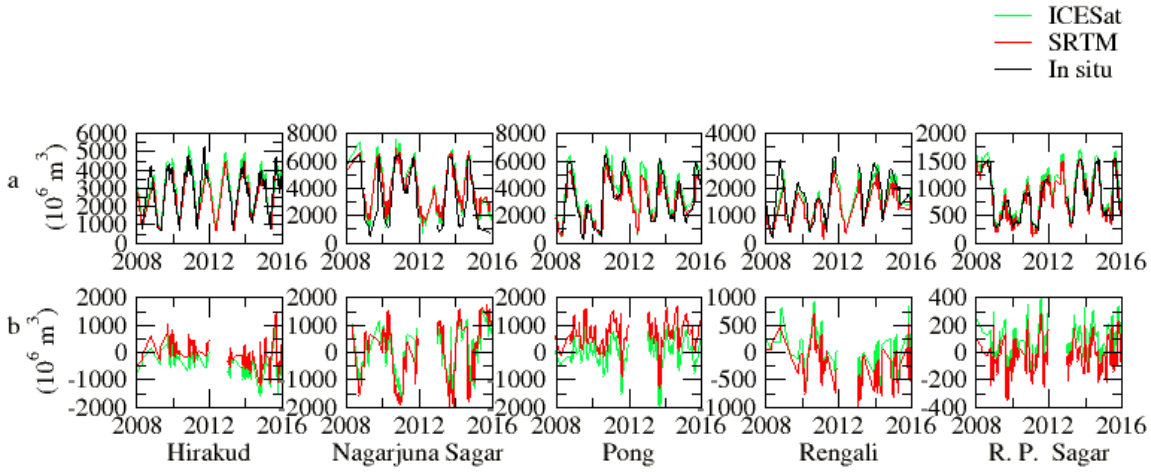


Figure 4.7 Validation results by comparing the remotely sensed storage with gauge observations. (a) is the comparison among absolute storage values; (b) is the comparison of storage difference (remotely sensed storage minus gauge data)

As shown in Figure 4.7a, both the MODIS-SRTM and MODIS-ICESat based approaches perform well overall. The time series from these two approaches closely match the gauge values for reservoir storage. In order to highlight the differences between the DEM and ICESat based approaches, Figure 4.7b compares the storage errors (against the gauge observations) from these two datasets. The error statistics are provided in Table 4.3. Among each of the five reservoirs, the NRMSE of the MODIS-

SRTM approach ranges from 18.14% to 40.06%, with a mean value of 25.24%. The relative bias values are from -12.25% to 20.32%. Meanwhile, the NRMSE of the MODIS-ICESat approach ranges from 14.20% to 24.60%, with a mean value of 19%. The bias values are from -8.62% to 3.75%. In terms of accuracy, MODIS-ICESat generally outperformed the DEM based approach—but the performances of two approaches are relatively close to each other. For the NJSagar reservoir, the NRMSE is 27.58% of the MODIS-SRTM approach and 24.6% of the ICESat based approach. For the R. P. Sagar reservoir, the DEM results are more accurate than the ICESat results. The NRMSE is 15.06%, which is 3.67% better than the ICESat based approach. For the Hirakud, Pong, and RP Sagar reservoirs, the MODIS-ICESat approach shows a superior accuracy when validating with the gauge data. The higher accuracy of the MODIS-ICESat approach is due to the higher vertical accuracy of ICESat elevation. In addition, the longer observation period of ICESat makes it capable of capturing a more comprehensive topography of the reservoirs, which results in a more accurate *A*-*H* relationship. In short, compared with the MODIS-ICESat approach, the MODIS-SRTM approach has a significant improvement with regard to spatial representation. However, the "cost" associated with monitoring more reservoirs (via MODIS-SRTM) is a bit of a loss of retrieval accuracy.

Table 4.3 Comparisons of the Validation Results from the MODIS-SRTM and MODIS-ICESat Approaches

		Hirakud	NJSagar	Pong	Rengali	R.P. Sagar
NRSME (%)	ICESat	14.20	24.60	17.79	19.71	18.73
	SRTM	18.14	27.58	27.76	23.87	15.62
Relative Bias (%)	ICESat	-1.87	3.75	0.97	-2.81	-8.62
	SRTM	-12.25	2.41	20.32	-13.51	-3.24

4.4.3. Uncertainty analysis

4.4.3.1. Uncertainty analysis due to the *A*-*H* relationship

The storage uncertainty caused by the *A*-*H* relationship error is analyzed. The *A*-*H* errors are primarily attributed to two sources: the use of partial bathymetry information to represent the *A*-*H* relationship for the entire reservoir, and the the limited precision of the DEM data.

Because the DEM dataset only represents the part of the bathymetry above the water surface when SRTM measurements were made, it was assumed that the unmeasured parts shared the same *A*-*H* relationship. To quantify the uncertainty associated with this assumption, we calculated the difference of storage by using the maximum (k_{max}) and minimum (k_{min}) slopes of the *A*-*H* relationship.

Three scenarios are used in Figure 4.8a to illustrate the process of quantifying the uncertainty from the first source. Figure 4.8a (I) shows a simplified sectional view of a reservoir when the terrain data was collected by the SRTM in 2000. The water surface area at that time is labeled as A_1, and the area of the reservoir bottom is A_2. k_{max} and k_{min}

are the maximum and minimum slope of the *A-H* relationship for the unknown part below the water surface (in theory). Because the uncertainty due to the first source can be quantified through k_{max} and k_{min}, the calculation of uncertainty is shown using equations (4.3) to (4.6).

The reservoir storage at the time of date collection is V_2 in equation (4.3).

$$V_2 = V_c - (A_c + A_1)(A_c - A_1)k/2 \quad (4.3)$$

We want to estimate the k_{max} and k_{min} and by giving the area of the reservoir bottom the two different assumptions shown in Figure 4.8 a (Ⅱ) and Figure 4.8 a (Ⅲ) via equations (4.4) and (4.5).

$$k_{\min} = \frac{V_2}{(A_1 + A_2)(A_1 - A_2)} = \frac{V_2}{A_1^2} \quad (4.4)$$

$$k_{\max} = \frac{V_2}{(A_1 + A_2)(A_1 - A_2)} = \frac{V_2}{(A_1 + A_{rs}^{\min})(A_1 - A_{rs}^{\min})} \quad (4.5)$$

In Figure 4.8 a (Ⅱ), the minimum value of A_2 (which is 0) is used to estimate the maximum value of k. In Figure 4.8 a (Ⅲ), the maximum value of A_2 (which is equal to the minimum water surface area from MODIS in the research period) is used to estimate the minimum value of *k*. After getting k_{max} and k_{min} from the above steps, the uncertainty due to the first source can be calculated through equation (4.6).

$$\Delta V = (A_c + A_{RS})(A_c - A_{RS})(k_{\max} - k_{\min})/2 \quad (4.6)$$

The precision of the DEM measurements also has an impact on the accuracy of the *A-H* relationship. We borrow the bias characteristic of SRTM from [Surazakov and Aizen, 2006], and consider it is related to the degree changes. We take Statistics of

SRTM DEM minus the DEM 1977 differences over glacier-free areas as a measure of surface change error (Δh).

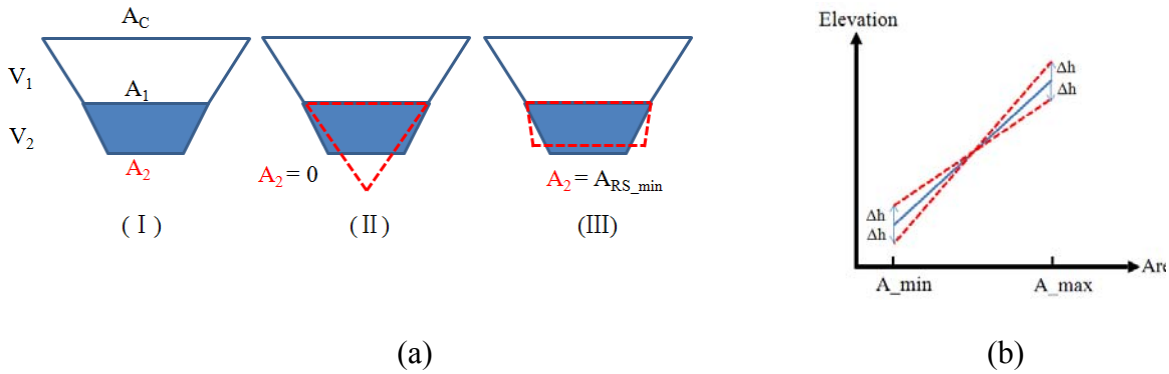
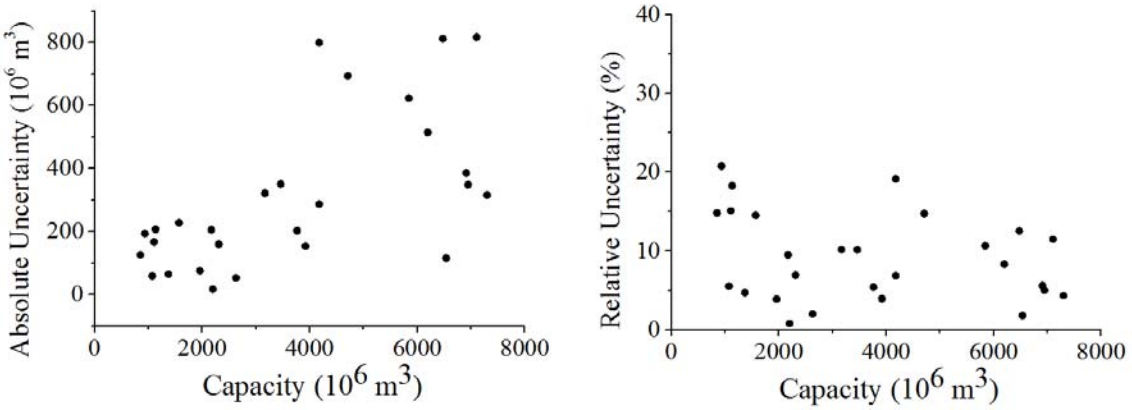


Figure 4.8 (a). Simplified example of quantifying the uncertainty caused by the unmeasured *A-H* relationship; (b) simplified example of quantifying the uncertainty caused by the SRTM DEM error

Based on the characteristics of each reservoir, the Δh is set to 0.5 m ~ 1m. The uncertainty due to the DEM error is calculated as shown in Figure 4.8b. The uncertainty for each reservoir is shown in Figure 4.9. For all 27 reservoirs, the absolute uncertainty due to the *A-H* relationship ranges from 53 to 812 km^3 (Figure 4.9a), with an average of 329 km^3. The absolute uncertainty has a positive relationship with the reservoir capacity. For every 1000 km^3 increase in reservoir capacity, the uncertainty will increase by 59 km^3 (based on the statistics from 27 reservoirs, $p<0.01$). The relative uncertainty due to the *A-H* relationship ranges from 0.77% to 27.12%, with a mean value of 9.47% (Figure 9b). There is no significant relationship detected between the relative uncertainty and the capacity. Among of these reservoirs, Matatila has the largest relative uncertainty of 27.12% (307 km^3)—while its capacity is only 1132 km^3. This is because the Matatila

reservoir has a steep slope, there is an assumption of a large DEM error when quantifying the uncertainty, which results in an unusually large relative uncertainty (if comparing with reservoirs which have a similar capacity).



(a) (b)

Figure 4.9 Uncertainty analysis results: (a) absolute uncertainty due to SRTM DEM; (b) relative uncertainty due to SRTM DEM.

4.4.3.2. Uncertainty analysis due to the mixture of pixels affecting water surface area estimation

In this study, an improved version of the k-means algorithm was used to estimate the surface water area for each reservoir. However, due to the limited spatial resolution of remote sensing data, a pixel that is classified into a land type may actually belong to multiple types. In order to quantify the bias caused by these "mixture pixels", a fuzzy clustering Fuzzy K-Means (FKM) [Bezdek, 2013] was utilized. Instead of assigning a certain type of land cover to each of the pixels, FKM uses "memberships" to denote the percentiles of each of the land types within one pixel.

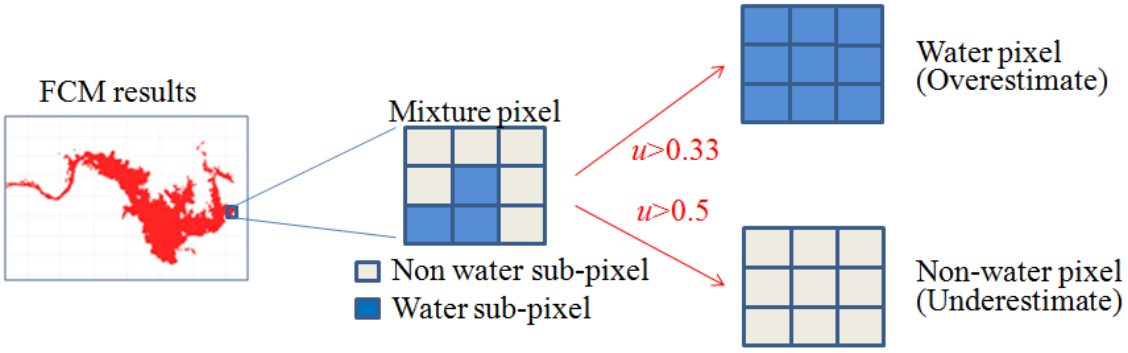
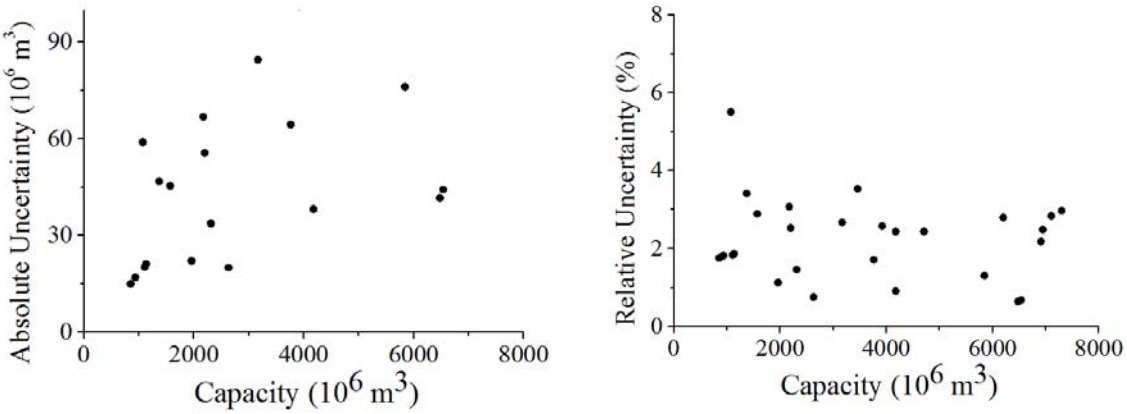


Figure 4.10 A simplified example showing the process for quantifying the uncertainty caused by mixture pixels

During the process of water surface estimation, two kinds of bias may be associated with the classification. One is the overestimation that occurs when the mixture pixels are classified as water. The other bias is the underestimation that occurs when mixture pixels were considered as non-water when estimating the surface water area. For evaluating the impact from neglecting the mixture pixels, we first applied the FKM algorithm to all pixels to get the fuzzy classification results. Each classification result is represented by membership from 0 to 1, which stands for the percentage of water in the individual pixel. Next two thresholds, 0.33 and 0.5, were chosen to de-fuzzy the classification results from FKM. Then the number of water pixels was counted, and the reservoir storage was calculated by using these two water surface areas. Finally, the reservoir storage uncertainty due to mixture pixels was evaluated by using the difference between the reservoir storage values.

The absolute and relative uncertainty is shown in Figures 4.11a and 4.11b. Because the mixture pixels are usually distributed along the shorelines of reservoirs—and the

large reservoir are likely to be associated with long shorelines—the uncertainty due to mixture pixels has a strong relationship with reservoir capacity. As demonstrated in Figure 4.11 a, the absolute uncertainty has a similar trend as those shown in Figure 4.9 a—but it is about an order of magnitude smaller. For relative uncertainty, there was no noticeable trend when the size of the reservoir changed. The relative uncertainty ranged from 0.76% to 5.5%, and its mean value was 2.23% (Figure 4.11 b).



(a) (b)

Figure 4.11 Uncertainty analysis results: (a) absolute uncertainty due to mixture pixels; (b) relative uncertainty due to mixture pixels.

4.5. Conclusions and discussions

In this study, a new method was developed to estimate reservoir storage variations with large spatial coverage by leveraging SRTM DEM data. Validation results against the gauge observations over 11 reservoirs in South Asia suggest that the remote storage values have a relatively high accuracy level (with R^2 values ranging from 0.47 to 0.91). By applying the newly developed algorithm to South Asia, a total of 27 reservoirs can be

monitored. The integrated storage capacity of these reservoirs is 118.76 km^3, which represent 45% of the overall storage in the region.

This method makes new contributions to reservoir remote sensing. It is the first time that SRTM DEM was used with other satellite data for establishing a reservoir storage data set which significantly extends the spatial coverage in South Asia. This is because the use of SRTM data has alleviated the limitation resulting from the sparseness of ICEsat tracks.

This algorithm still has some limitations which need to be discussed. Firstly, the accuracy of the proposed algorithm depends on the water level when the DEM data was collected. For certain reservoirs which had relatively high water levels when the surface elevation data were collected by SRTM, the *A*-*H* relationships (using the current approach) would not be sufficient to represent the entire reservoir. In these cases, the performance of the algorithm will be limited. Secondly, because the performance of this algorithm depends on the accuracy of the DEM data (which is not as good as ICESat), the overall accuracy of the DEM based approach is smaller than the ICESat based approach. Nonetheless, the benefits of the extended number of reservoirs outweigh the constraints. In order to better evaluate the impact from these limitations, we quantify the uncertainty associated with the storage estimations from *A*-*H* relationship (which is essentially inferred from the SRTM DEM).

This algorithm can result in reservoir storage products which support water management at a large scale. For instance, given the long term availability of high spatial resolution satellite, this approach can monitor much smaller size reservoirs. This

algorithm may also contribute to future satellite missions such as SWOT , which will provide a direct water surface measurement for about two-thirds of global lake and reservoir storage (with an area > 0.06 km^2).

CHAPTER V

SUMMARY

5.1. Conclusions

Satellite remote sensing has offered a unique opportunity to study the Earth from space. Its global coverage (which is free of geographical limitations) has shed new light on flood monitoring and forecasting in these international river basins. By combining the information about water surface extent and water surface elevation, many studies in remote sensing have been carried out for estimating reservoir storage variations. However, simultaneously optimizing the spatial and temporal resolution of satellite data remains the biggest challenge towards monitoring more reservoirs with high accuracy. The dissertation research presented in Chapters II-IV primarily focuses on improving the spatial and temporal representation and retrieval accuracy of reservoir storage from satellites. In short the overarching goal of this study is to generate a remotely sensed reservoir storage data set over the South Asia region with high temporal resolution and large spatial coverage.

Aiming at first generating a reservoir storage product for South Asia with high accuracy, in Chapter II, an algorithm was developed by using MODIS and ICESat data. In order to improve the accuracy of water surface area estimations for relatively small reservoirs, a novel classification algorithm was developed. In this study, storage information was retrieved for a total of 21 reservoirs, which represents 28% of the integrated reservoir capacity in South Asia. The satellite-based reservoir elevation and

storage were validated by gauge observations over five reservoirs. The storage estimates were highly correlated with observations (i.e., correlation coefficients larger than 0.9), with normalized root mean square error (NRMSE) ranging from 9.51% to 25.20%.

MODIS-based reservoir storage monitoring approaches have been constrained to work under cloud-free days, which significantly limits the applications of satellite-based estimates to a variety of disciplines and areas. Chapter III provides a new integrated technique of a continuous monitoring capability of reservoir storage by leveraging cloud-penetrating passive microwave observations. Validation results against *in situ* observations over the four testing reservoirs in South Asia suggest that the filtered AMSR-E 4-day storage estimations have a relatively high accuracy level (with R^2 values ranging from 0.41 to 0.74).

In Chapter IV, for extending the spatial cover of the existing reservoir data set, a new method was developed to estimate reservoir storage variations with large spatial coverage by leveraging the SRTM DEM. Validation results against the gauge observations over 11 reservoirs in South Asia suggest that the remote storage values have a relatively high accuracy level (with R^2 values ranging from 0.47 to 0.91). By applying the newly developed algorithm to South Asia, a total of 27 reservoirs can be monitored. The integrated storage capacity of these reservoirs is 118.76 km^3, which represent 45% of the overall storage in the region.

Overall, accuracy of reservoir storage retrievals from satellite remote sensing have been improved due to improvements in water surface estimation from MODIS and the use of water surface height from ICESat. In addition, the temporal resolution and spatial

representation have been strengthened by using microwave passive data and DEM, respectively.

5.2. Recommendations of future research

(1) This algorithm can result in reservoir storage products which support water management in long term (climatological) observations and short term monitoring. For instance, given the long term availability of passive microwave remote sensing data—such as that from the Special Sensor Microwave Imager (SSM/I) and the Special Sensor Microwave Imager Sounder (SSMIS) on board the Defense Meteorological Satellite Program (DMSP) satellites—a climatological reservoir record can be generated.

(2) Even though the elevation information extracted from SRTM would make the product with extended spatial coverage, the retrieval accuracy is decreased by the limitations of SRMT. With the launch of ICESat2 mission in 2016 and the SWOT mission in 2021, a significantly greater number of reservoirs will be able to be studied at much higher spatial resolutions with higher quality.

REFERENCES

Adhikari, P., Y. Hong, K. R. Douglas, D. B. Kirschbaum, J. Gourley, R. Adler and G. R. Brakenridge (2010). A digitized global flood inventory (1998–2008): compilation and preliminary results. *Natural Hazards*, 55, 405-422.

Akay, B. and D. Karaboga (2012). A modified artificial bee colony algorithm for real-parameter optimization. *Information Sciences*, 192, 120-142.

Alsdorf, D., C. Birkett, T. Dunne, J. Melack and L. Hess (2001). Water level changes in a large Amazon lake measured with spaceborne radar interferometry and altimetry. *Geophysical Research Letters*, 28, 2671-2674.

Alsdorf, D. E., E. Rodríguez and D. P. Lettenmaier (2007). Measuring surface water from space. *Reviews of Geophysics*, 45, RG2002.

Bai, T., L. Wu, J. Chang and Q. Huang (2015). Multi-objective optimal operation model of cascade reservoirs and its application on water and sediment regulation. *Water Resources Management*, 29, 2751-2770.

Batt, R. D. and S. R. Carpenter (2012). Free‐water lake metabolism: addressing noisy time series with a Kalman filter. *Limnology and Oceanography: Methods*, 10, 20-30.

Berthier, E., Y. Arnaud, C. Vincent and F. Remy (2006). Biases of SRTM in high‐mountain areas: Implications for the monitoring of glacier volume changes. *Geophysical Research Letters*, 33, L08502.

Bezdek, J. C. (2013). Pattern recognition with fuzzy objective function algorithms. *Springer Science & Business Media*, Berlin, Germany.

Biancamaria, S., F. Hossain and D. Lettenmaier (2011). Forecasting transboundary river water elevations from space. *Geophysical Research Letters*, 38, L11401.

Birkett, C. M. and B. Beckley (2010). Investigating the performance of the Jason-2/OSTM radar altimeter over lakes and reservoirs. *Marine Geodesy*, 33, 204-238.

Brakenridge, G. R., S. V. Nghiem, E. Anderson and R. Mic (2007). Orbital microwave measurement of river discharge and ice status. *Water Resources Research*, 43, W04405.

Brest, J., A. Zamuda, B. Bošković, M. S. Maučec and V. Žumer (2008). High-dimensional real-parameter optimization using self-adaptive differential evolution algorithm with population size reduction. *IEEE Congress on Evolutionary Computation*, 2032-2039, Hong Kong, China.

Calmant, S., F. Seyler and J. F. Cretaux (2008). Monitoring continental surface waters by satellite altimetry. *Surveys in geophysics*, 29, 247-269.

Crétaux, J.-F. and C. Birkett (2006). Lake studies from satellite radar altimetry. *Comptes Rendus Geoscience*, 338, 1098-1112.

Crétaux, J.-F., W. Jelinski, S. Calmant, A. Kouraev, V. Vuglinski, M. Bergé-Nguyen, M.-C. Gennero, F. Nino, R. A. Del Rio and A. Cazenave (2011). SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Advances in Space Research*, 47, 1497-1507.

Crow, W. and M. Van den Berg (2010). An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research*, 46, W12519.

Draper, C. S., J. P. Walker, P. J. Steinle, R. A. de Jeu and T. R. Holmes (2009). An evaluation of AMSR–E derived soil moisture over Australia. *Remote Sensing of Environment*, 113, 703-710.

Duan, Z. and W. G. M. Bastiaanssen (2013). Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. *Remote Sensing of Environment*, 134, 403-416.

Fallah-Mehdipour, E., O. B. Haddad and M. Mariño (2012). Real-time operation of reservoir system by genetic programming. *Water Resources Management*, 26, 4091-4103.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez and L. Roth (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45, RG2004.

Fily, M., A. Royer, K. Goıta and C. Prigent (2003). A simple retrieval method for land surface temperature and fraction of water surface determination from satellite microwave brightness temperatures in sub-arctic areas. *Remote Sensing of Environment*, 85, 328-338.

Fluet-Chouinard, E., B. Lehner, L.-M. Rebelo, F. Papa and S. K. Hamilton (2015). Development of a global inundation map at high spatial resolution from

topographic downscaling of coarse-scale remote sensing data. *Remote Sensing of Environment*, 158, 348-361.

Gao, H. (2015). Satellite remote sensing of large lakes and reservoirs: From elevation and area to storage. *Wiley Interdisciplinary Reviews: Water*, 2, 147-157.

Gao, H., C. Birkett and D. P. Lettenmaier (2012). Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, 48, W09504.

Gao, H., R. Fu, R. E. Dickinson and R. I. N. Juarez (2008). A practical method for retrieving land surface temperature from AMSR-E over the Amazon Forest. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 193-199.

Gao, H., E. Wood, T. Jackson, M. Drusch and R. Bindlish (2006). Using TRMM/TMI to retrieve surface soil moisture over the southern United States from 1998 to 2002. *Journal of Hydrometeorology*, 7, 23-38.

Gao, H., S. Zhang, M. Durand and H. Lee (2016). Satellite Remote Sensing of Lakes and Wetlands. *Hydrologic Remote Sensing: Capacity Building for Sustainability and Resilience*, Chapter 4, Taylor & Francis Group, FL, United States.

Gelb, A. (1974). *Applied optimal estimation*: MIT press, MA, United States.

Haddeland, I., J. Heinke, H. Biemans, S. Eisner, M. Flörke, N. Hanasaki, M. Konzmann, F. Ludwig, Y. Masaki and J. Schewe (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111, 3251-3256.

Hejazi, H., H. Mohabati, S. Hosseinian and M. Abedi (2011). Differential evolution algorithm for security-constrained energy and reserve optimization considering credible contingencies. *IEEE Transactions on Power Systems*, 26, 1145-1155.

Holmes, T., R. De Jeu, M. Owe and A. Dolman (2009). Land surface temperature from Ka band (37 GHz) passive microwave observations. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114, D04113.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman and E. F. Stocker (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8, 38-55.

Islam, A., S. Bala and M. Haque (2010). Flood inundation map of Bangladesh using MODIS time‐series images. *Journal of Flood Risk Management*, 3, 210-222.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31, 651-666.

Ji, L., L. Zhang and B. Wylie (2009). Analysis of dynamic thresholds for the normalized difference water index. *Photogrammetric Engineering & Remote Sensing*, 75, 1307-1317.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35-45.

Knowles, K., M. Savoie, R. Armstrong and M. Brodzik (2006). AMSR-E/Aqua Daily EASE-Grid Brightness Temperatures. *Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center.*

Lehner, B., C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken and J. Magome (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9, 494-502.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan and E. F. Wood (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51, 7309-7342.

Li, J., H. Fang and L. Yang (2011). Mapping Lake Level Changes using ICESat/GLAS Satellite Laser Altimetry Data–A Case Study in Arid Regions of Central Asia. *Proceedings of SPIE*, MIPPR 2011: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications, 80060J.

Liebe, J., N. Van De Giesen and M. Andreini (2005). Estimation of small reservoir storage capacities in a semi-arid environment: A case study in the Upper East Region of Ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, 30, 448-454.

Mateo, C. M., N. Hanasaki, D. Komori, K. Tanaka, M. Kiguchi, A. Champathong, T. Sukhapunnaphan, D. Yamazaki and T. Oki (2014). Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models. *Water Resources Research*, 50, 7245-7266.

Maulik, U. and I. Saha (2010). Automatic fuzzy clustering using modified differential evolution for image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 3503-3510.

McKellip, R., B. Beckley, C. B. S. Birkett, B. Doorn, B. Grant, L. Estep, R. Moore, K. Morris, K. Ross and G. Terrie (2004). *PECAD's global reservoir and lake monitor: A systems engineering report*, version 1.0, NASA/John C. Stennis Space Center.

Pan, M. and E. F. Wood (2006). Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *Journal of Hydrometeorology*, 7, 534-547.

Papa, F., F. Frappart, A. Güntner, C. Prigent, F. Aires, A. C. Getirana and R. Maurer (2013). Surface freshwater storage and variability in the Amazon basin from multi‐satellite observations, 1993‐2007. *Journal of Geophysical Research: Atmospheres*, 118, 951- 965.

Papa, F., C. Prigent, F. Aires, C. Jimenez, W. Rossow and E. Matthews (2010). Interannual variability of surface water extent at the global scale, 1993–2004. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115, D12111.

Papa, F., C. Prigent, F. Durand and W. Rossow (2006). Wetland dynamics using a suite of satellite observations: A case study of application and evaluation for the Indian Subcontinent. *Geophysical Research Letters*, 33, L08401.

Phan, V. H., R. Lindenbergh and M. Menenti (2012). ICESat derived elevation changes of Tibetan lakes between 2003 and 2009. *International Journal of Applied Earth Observation and Geoinformation*, 17, 12-22.

Pulwarty, R. S. and M. V. Sivakumar (2014). Information systems in a changing climate: Early warnings and drought risk management. *Weather and Climate Extremes*, 3, 14-21.

Rabus, B., M. Eineder, A. Roth and R. Bamler (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57, 241-262.

Reichle, R. H., D. B. McLaughlin and D. Entekhabi (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130, 103-114.

Rekanos, I. T. (2008). Shape reconstruction of a perfectly conducting scatterer using differential evolution and particle swarm optimization. *IEEE Transactions on Geoscience and Remote sensing*, 46, 1967-1974.

Rocca, P., G. Oliveri and A. Massa (2011). Differential evolution as applied to electromagnetics. *IEEE Antennas and Propagation Magazine*, 53, 38-49.

Rodrigues, L. N., E. E. Sano, T. S. Steenhuis and D. P. Passo (2012). Estimation of Small Reservoir Storage Capacities with Remote Sensing in the Brazilian Savannah Region. *Water Resources Management*, 26, 873-882.

Ronkkonen, J., S. Kukkonen and K. V. Price (2005). Real-parameter optimization with differential evolution. *Proc. IEEE CEC*, 506-513.

Schutz, B., H. Zwally, C. Shuman, D. Hancock and J. DiMarzio (2005). Overview of the ICESat mission. *Geophysical Research Letters*, 32, L21S01.

Schwatke, C., D. Dettmering, W. Bosch and F. Seitz (2015). DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, 19, 4345-4364.

Shuman, C., H. Zwally, B. Schutz, A. Brenner, J. DiMarzio, V. Suchdeo and H. Fricker (2006). ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. *Geophysical Research Letters*, 33, L07501.

Smith, B. E., H. A. Fricker, I. R. Joughin and S. Tulaczyk (2009). An inventory of active subglacial lakes in Antarctica detected by ICESat (2003–2008). *Journal of Glaciology*, 55, 573-595.

Smith, L. C. and T. M. Pavelsky (2009). Remote sensing of volumetric storage changes in lakes. *Earth Surface Processes and Landforms*, 34, 1353-1358.

Song, C., B. Huang and L. Ke (2013). Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. *Remote Sensing of Environment*, 135, 25-35.

Storn, R. and K. Price (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341-359.

Surazakov, A. B. and V. B. Aizen (2006). Estimating volume change of mountain glaciers using SRTM and map-based topographic data. *IEEE Transactions on Geoscience and Remote sensing*, 44, 2991-2995.

Vicente-Serrano, S. M., S. Beguería, L. Gimeno, L. Eklundh, G. Giuliani, D. Weston, A. El Kenawy, J. I. López-Moreno, R. Nieto and T. Ayenew (2012). Challenges for drought mitigation in Africa: The potential use of geospatial data and drought information systems. *Applied Geography*, 34, 471-486.

Wang, X., P. Gong, Y. Zhao, Y. Xu, X. Cheng, Z. Niu, Z. Luo, H. Huang, F. Sun and X. Li (2013). Water-level changes in China's large lakes determined from ICESat/GLAS data. *Remote Sensing of Environment*, 132, 131-144.

Watts, J. D., J. S. Kimball, L. A. Jones, R. Schroeder and K. C. McDonald (2012). Satellite Microwave remote sensing of contrasting surface water inundation changes within the Arctic–Boreal Region. *Remote Sensing of Environment*, 127, 223-236.

Wolf, A. T., J. A. Natharius, J. J. Danielson, B. S. Ward and J. K. Pender (1999). International river basins of the world. *International Journal of Water Resources Development*, 15, 387-427.

Wu, H., R. F. Adler, Y. Hong, Y. Tian and F. Policelli (2012). Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *Journal of Hydrometeorology*, 13, 1268-1284.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 3025-3033.

Yildiz, A. R. (2013). A new hybrid differential evolution algorithm for the selection of optimal machining parameters in milling operations. *Applied Soft Computing*, 13, 1561-1566.

Zhang, G., J. Cheng, M. Gheorghe and Q. Meng (2013). A hybrid approach based on differential evolution and tissue membrane systems for solving constrained

manufacturing parameter optimization problems. *Applied Soft Computing*, 13, 1528-1542.

Zhang, G., H. Xie, S. Kang, D. Yi and S. F. Ackley (2011). Monitoring lake level changes on the Tibetan Plateau using ICESat altimetry data (2003–2009). *Remote Sensing of Environment*, 115, 1733-1742.

Zhang, S. and H. Gao (2016). A novel algorithm for monitoring reservoirs under all‐weather conditions at a high temporal resolution through passive microwave remote sensing. *Geophysical Research Letters*, 43, 8052-8059.

Zhang, S., H. Gao and B. S. Naz (2014). Monitoring reservoir storage in South Asia from multisatellite remote sensing. *Water Resources Research*, 50, 8927-8943.

Zhao, G., H. Gao, B. S. Naz, S.-C. Kao and N. Voisin (2016). Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources*, 98, 16-31.

Zhong, Y., S. Zhang and L. Zhang (2013). Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6, 2290-2301.

Zhou, T., B. Nijssen, H. Gao and D. P. Lettenmaier (2016). The contribution of reservoirs to global land surface water storage variations. *Journal of Hydrometeorology*, 17, 309-325.

Zwally, H., B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock and D. Harding (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34, 405-445.

Zwally, H. J., D. Yi, R. Kwok and Y. Zhao (2008). ICESat measurements of sea ice freeboard and estimates of sea ice thickness in the Weddell Sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 113, C02S15.

APPENDIX

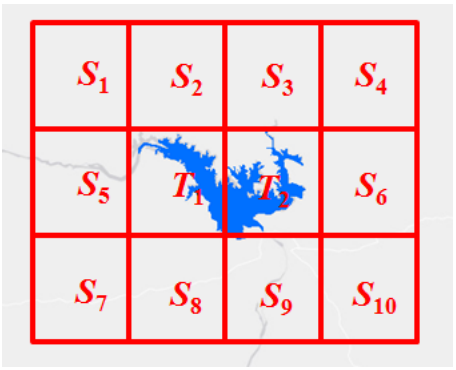
Introduction

This supplemental information document contains general information of the reservoirs in this study (Table A1), pixel selections for each reservoir (Figure A1), detailed information about optimization of the weight coefficients via Differential Evolution, and Kalman Filtering based noise reduction.

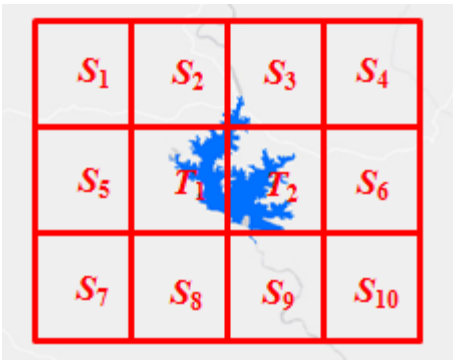
Table A1. General information of the reservoirs in this study

Reservoir	Hirakud	Nagarjuna Sagar	Pong	Rengali
Lat (ºN), lon (ºE)	21.52, 83.85	16.57, 79.31	31.97, 75.95	21.28, 85.03
Capacity (10^6 m^3)	4709	6538	6946	3168
Area at capacity (km^2)	603	285	260	392
Elevation at capacity (a.s.l. m)	192.02	179.83	426.72	123.44
River	Mahanadi River	Krishna river	Beas River	Brahmani River
*Purpose	I, E	I, E	I, E	I
NRMSE	12.77%	24.75%	9.51%	25.20%

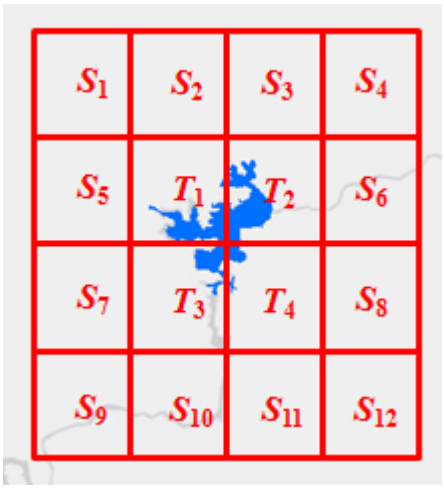
*I is irrigation, E is electricity generation



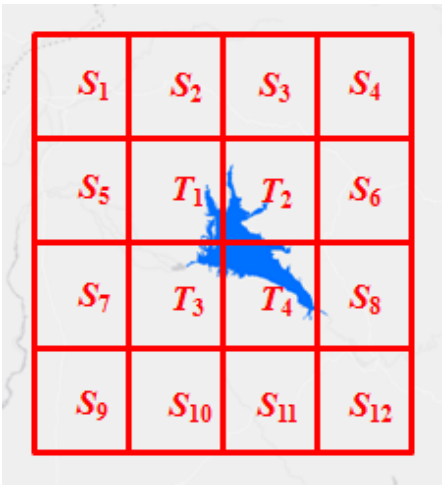
(a) Hirakud



(c) Rengali



(b) Nagarjuna Sagar



(d) Pong

Figure A1. The selection of target and surrounding pixels for (a) Hirakud; (b) Nagarjuna Sagar; (c) Renglai; and (d) Pong reservoir.

Optimization of the weight coefficients via the Differential Evolution (DE) algorithm

The DE algorithm is a floating-point encoded evolutionary algorithm which aims to find the optimal solution to maximize or minimize the objective function [Storn and Price, 1997]. It has been applied widely to parameter optimization because of its efficiency in searching for and obtaining optimal parameters [Akay and Karaboga, 2012; Brest et al., 2008; Ronkkonen et al., 2005; Yildiz, 2013; Zhang et al., 2013]. DE generally consists of 5 steps: population initialization, mutation, crossover, selection, and evolution termination.

1) Population initialization. A population of *NP* individuals—each representing a search point in the space of feasible solutions, *G*—is initialized as $G = \{G_1, \cdots G_i \cdots, G_{NP}\}$. Because there is no prior knowledge about the solution space in this study, each individual is randomly generated—i.e. $G_i = \{ G_{i,1}, \cdots G_{i,j}, \cdots G_{i,d} \}$, where *d* is a number identifying the vector's dimension. In this study, each G_i stands for a possible combination of weight coefficients, and *d* is equal to the total number of weight coefficients (or the total number of target pixels and neighboring pixels). Here, *NP* is set to 30.

2) Mutation. For each individual G_i in the current generation, a corresponding mutant individual v_i can be generated using equation (A1):

$$v_i = G_{r_1} + F \cdot (G_{r_2} - G_{r_3}) \tag{A1}$$

where F is a scaling factor; G_{r_1}, G_{r_2} and G_{r_3} represent three individuals picked randomly from the generation; and r_1, r_2 and r_3 are random integers from 1 to NP (and $r_1 \neq r_2 \neq r_3$). The scaling factor F must be within the range of 0 to 2, and a value of 0.3 was selected in this study.

3) Crossover. To improve the diversity of the population, a crossover operation is applied after the mutation process. A trial individual, $Q_i = \{ q_{i,1}, \cdots q_{i,j} \cdots, q_{i,d} \}$, can be formed by exchanging the components of the mutant individual v_i and the target individual G_i. The crossover formula is described in equation (A2):

$$q_{i,j} = \begin{cases} v_{i,j}, & if \ \ rand(0,1) \le CR \ or \ \ j = rank(j) \\ & j = 1,2,...,d \\ G_{i,j}, & if \ \ rand(0,1) > CR \ or \ \ j \ne rank(j) \end{cases} \qquad \text{(A2)}$$

where $CR \in [0,1]$ is the crossover rate, $rand(0,1) \in [0,1]$ is a uniformly distributed random number, and *rank(j)* is a random integer from 0 to d. This operation can ensure that at least one pair of components from the mutant individual and a target individual (with the same rank) is exchanged, and (hence) the potential diversity of the population is enhanced. The value for CR was set at 0.6 in this study.

4) Selection. To keep the size of the population constant—and to decide which individual should be involved in the next generation—a selection was made between the target individual G_i and the trial individual Q_i, using equation (A3):

$$G_i^{'} = \begin{cases} Q_i, & if \ \ f(Q_i) \ge f(G_i) \\ G_i, & if \ \ f(Q_i) < f(G_i) \end{cases} \qquad \text{(A3)}$$

here *f*() is the objective function to be maximized/minimized, and $G_i^{'}$ is the *i*-th individual in the next generation.

5) Evolution termination. The number of iterations can either be set based on experience or by setting a threshold difference between two iterations as the stopping condition. Here, we set a fixed iteration number of 2000 as the stopping condition. The process will be iterated (from Step 2 to Step 5) until the stopping condition is met, and the optimal individual is produced.

Kalman Filter based noise reduction

In this study, the Kalman Filtering approach [Kalman, 1960] was used to reduce the high frequency noise contained in the remotely sensed water storage. The overall idea is to optimally estimate the reservoir storage by combining the noisy AMSR-E based results with the storage estimated from an annual cycle model. Created using the training dataset, the annual cycle model is a generalization of the seasonal storage climatology, which does not reflect interannual variations (i.e., results during wet years are the same as those during dry years). In contrast, the AMSR-E based 4-day reservoir storage data (obtained from Section 3.2.2.2) has valuable information buried in the high frequency noise. Thus, while both the model predicted storage and the AMSR-E based storage have uncertainties, the characteristics of these uncertainties are different. Through Kalman Filtering, the updated storage will have an optimal value (and uncertainty) that is better than that from either alone. The detailed process for reducing the noise of the AMSR-E based storage using a Kalman Filter is described in the following steps.

Step 1. Construction of the modeled and the AMSR-E based storage in accordance with the framework of the Kalman Filter

An annual cycle model was constructed from the MODIS based 16-day reservoir storage training dataset. The data points—on the same date but from different years (from 2000 to 2012, if/as available)—were averaged to represent the annual storage cycle of the reservoir of interest. Although the training dataset is nominally 16-day, it may not contain values for every 16-day interval due to cloud contamination. In all cases, however, a 4-day interval annual cycle was derived using linear interpolation between

the two nearest data points, to match the temporal resolution of the AMSR-E based results. This annual cycle was then repeated to tally the multi-year AMSR-E based storage estimations.

In order to use the Kalman Filter for noise reduction, the modeled and AMSR-E based storage values (as well as their uncertainties) needed to be expressed as a suit of matrices in accordance with the Kalman Filter framework. The basic assumption of these matrices is that the true storage T at time k is evolved from the storage at $(k-1)$ through equation (A4):

$$T(k) = A(k)T(k-1) + W \qquad \text{(A4)}$$

where $A(k)$ is the ratio between the modeled storage values from the annual cycle model at times k and k-1, and W is the uncertainty. The uncertainty W is assumed to be Gaussian Noise, with a zero mean value and a variance of w^2 (i.e., $W \sim (0, w^2)$). The value of w^2 is calculated as the mean square difference between the two adjacent storage values.

The relationship between the true reservoir storage $T(k)$ and the AMSR-E based storage $Z(k)$ is shown in equation (A5):

$$Z(k) = H\,T(k) + V \quad \text{(A5)}$$

where H is an identity matrix, and V is the uncertainty of $Z(k)$. For this study, H is equal to 1. Similar to W, the uncertainty V is assumed to be Gaussian Noise, with a zero mean value and a variance of v^2. The value of v^2 is the mean square difference between the two adjacent values of $Z(k)$ during the calibration period.

Step 2. Noise reduction based on Kalman Filtering

The optimal storage at time k after noise reduction, $X(k|k)$, can be calculated according equation (A6):

$$X(k \mid k) = X(k \mid k-1) + Kg(k)(Z(k) - H\,X(k \mid k-1)) \quad \text{(A6)}$$

where $Kg(k)$ is the Kalman gain, and $X(k|k\text{-}1)$ is the storage value at k. $X(k|k\text{-}1)$ is predicted from the optimal value at k-1 using equation (A7):

$$X(k \mid k-1) = A(k)\,X(k-1 \mid k-1) \quad \text{(A7)}$$

The Kalman gain, $Kg(k)$, is calculated from equation (A8):

$$Kg(k) = \frac{P(k \mid k-1)H^T}{H\,P(k \mid k-1)H^T + v^2} \qquad \text{(A8)}$$

where $P(k|k\text{-}1)$ is the error covariance of $X(k|k\text{-}1)$, as defined in equation (A9):

$$P(k \mid k-1) = A(k)P(k-1 \mid k-1)A^T(k) + w^2 \quad \text{(A9)}$$

By updating $P(k|k)$ using equation (A10), the Kalman filter can run recursively (for each additional time step k+1) through the equations (A6)-(A9).

$$P(k \mid k) = (I - Kg(k)H)P(k \mid k-1) \quad \text{(A10)}$$